WORLD POLITICS

A Quarterly Journal of International Relations

Volume 43
October 1990–July 1991

UNDER THE EDITORIAL SPONSORSHIP OF
CENTER OF INTERNATIONAL STUDIES
PRINCETON UNIVERSITY
PUBLISHED BY THE JOHNS HOPKINS UNIVERSITY PRESS

The editors invite submission of articles (up to 10,000 words, including footnotes) bearing upon problems in international relations and comparative politics. Manuscripts and footnotes should be doublespaced throughout and submitted in triplicate, together with an abstract of approximately 125 words. Footnotes should only cite sources; they should be numbered consecutively and placed at the end of the manuscript. Address: WORLD POLITICS, Bendheim Hall, Princeton, NJ 08544.

Authors may expect to be notified of the editors' decision within three months from date of submission. Procedures for reviewing manuscripts are based on the anonymity of the author and the confidentiality of readers' and editors' reports. Editorial policy does not provide for transmitting readers' evaluations to authors of rejected manuscripts. Exceptions may be made if, in the opinion of the editors, such comments contain substantive suggestions for improvement. WORLD POLITICS does not accept manuscripts that have already been published, are scheduled for publication elsewhere, or have been simultaneously submitted to another journal. Statements of fact and opinion appearing in the journal are made on the responsibility of the authors alone and do not imply the endorsement of the editors or publishers. The journal does not publish communications to the editor or rejoinders to specific articles. Scholars who believe they have been challenged are encouraged to submit an article that will advance the scholarly debate.

Reprint permission: Permission to reproduce or translate material should be sought from Permissions Department, The Johns Hopkins University Press, 701 W. 40th Street, Suite 275, Baltimore, MD 21211-2190.

Subscriptions: WORLD POLITICS, The John Hopkins University Press, Journals Publishing Division, 701 W. 40th Street, Suite 275, Baltimore, MD 21211-2190.
Individuals: $20 per year.
Institutions: $40 per year.
Single Copies: individuals, $7.00; institutions, $12.00.
For postage and handling in Canada and Mexico, please add $4.75 per year; outside of North America, $8.50.

Back Issues: WORLD POLITICS is available in microfilm or xerographic copies from University Microfilms, Ann Arbor, MI 48106. Two years after publication, complete volumes are available in microfiche form from KTO Microform, Route 100, Millwood, N.Y. 10546. Individual issues in the current and two preceding volumes may be ordered from The Johns Hopkins University Press, Journals Publishing Division, 701 W. 40th Street, Suite 275, Baltimore, MD 21211-2190.

Advertising: Journals Advertising Coordinator, The Johns Hopkins University Press, Journals Publishing Division, 701 W. 40th Street, Suite 275, Baltimore, MD 21211-2190.

WORLD POLITICS is indexed by *ABC POL SCI, Book Review Index, Combined Retrospective Index Sets (CRIS), Combined Retrospective Index to Book Reviews in Scholarly Journals, 1886–1974 (RSJ), PAIS Bulletin,* and *Social Science Index*; it is abstracted and indexed in *United States Political Science Documents, International Bibliography of Periodical Literature (IBZ), International Bibliography of Book Reviews (IBR)*, and *Current Military and Political Literature*. Abstracts of articles also appear in *Book Review Digest, Historical Abstracts, International Development Abstracts, International Political Science Abstracts*, and *Political Science Abstracts*.

WORLD POLITICS (ISSN 0043-8871). Published quarterly by The Johns Hopkins University Press. Vol. 43, October 1990–July 1991. Second class postage paid at Baltimore, MD, and additional mailing offices. POSTMASTER: Send address changes to WORLD POLITICS, The Johns Hopkins University Press, Journals Publishing Division, 701 W. 40th Street, Suite 275, Baltimore, MD 21211-2190. Printed in the United States of America by Princeton University Press.

CONTENTS

No. 1—October 1990

No. 2—January 1991

No. 3—April 1991

No. 4—July 1991

WORLD POLITICS

Vol. 43 • October 1990 • No. 1

CONTENTS

THE CONTRIBUTORS

YOUNG-KWAN YOON is Assistant Professor in the Department of International Relations at the College of Social Science, Seoul National University, Seoul, Korea. His principal interests are foreign investments and industry-finance relations among the economic powers and the political economy of East Asian development.

BRUCE BUENO DE MESQUITA is a Senior Fellow at the Hoover Institution, Stanford University, and Professor of Political Science at the University of Rochester. The author of *The War Trap* (1981) and (together with David Newman and Alvin Rabushka) *Forecasting Political Events: Hong Kong's Future* (1985), he is currently completing *War and Reason* (with David Lalman), which is forthcoming from Yale University Press.

TYRENE WHITE is Assistant Professor of Political Science at Swarthmore College. She has written on rural politics and family planning in China and is currently completing a book manuscript entitled *Against the Grain: The State-Peasant Struggle over Child-Bearing in Contemporary China.*

ZEEV MAOZ is Associate Professor of Political Science at the University of Haifa, Israel. He has also taught at New York University. He is the winner of the 1989 Karl W. Deutsch Award of the International Studies Association. His most recent books are *National Choices and International Processes* (1990) and *Paradoxes of War: On the Art of National Self-Entrapment* (1989).

RUSSELL MARDON is Assistant Professor of Political Science at California State University, Fresno.

JOHN S. ODELL is Professor of International Relations and Director of the Center for International Studies at the University of Southern California. He is the author of *U.S. International Monetary Policy: Markets, Power, and Ideas as Sources of Change* (1982), coauthor of *Anti-Protection: Changing Forces in U.S. Trade Politics* (1987), and coeditor of *International Trade Policies: Gains from Exchange between Economics and Political Science* (1990).

WORLD POLITICS

Vol. 43 • January 1991 • No. 2

CONTENTS

THE CONTRIBUTORS

JAMES D. FEARON is a Ph.D. candidate in Political Science at the University of California, Berkeley. His dissertation examines the question of when threats used in international crises deter and when they provoke.

PHILIP G. ROEDER is Assistant Professor of Political Science at the University of California, San Diego. He is the author of *Soviet Political Dynamics: Development of the First Leninist Polity* (1988) and is currently completing a book that analyzes how institutions created by the Stalin constitution have shaped and limited reform of Soviet politics.

STEVEN R. DAVID is Associate Professor of Political Science at The Johns Hopkins University. He has written widely on the Third World, with an emphasis on security issues. This article is derived from a forthcoming book, *Choosing Sides: Third World Alignment and Realignment.*

MOHAMMED AYOOB is Associate Professor of International Relations, James Madison College, Michigan State University. During 1989–90 he was an SSRC-MacArthur Foundation Fellow in International Peace and Security at the Center of International Studies, Princeton University. His recent publications include *India and Southeast Asia: Indian Perceptions and Policies* (1990) and two edited volumes entitled *Leadership Perceptions and National Security: The Southeast Asian Experience* (1989) and *Regional Security in the Third World* (1986). He is currently working on a project exploring the security problematic of Third World states.

JANINE LUDLAM has done graduate work in Russian Area Studies at Georgetown University and is an observer of Soviet domestic politics currently residing in Moscow.

WORLD POLITICS

Vol. 43 • April 1991 • No. 3

CONTENTS

THE CONTRIBUTORS

BARRY NALEBUFF is Professor of Economics and Management at Yale University's School of Organization and Management. He is the author of *Thinking Strategically: A Competitive Edge in Business, Politics, and Everyday Life* (1991).

STEPHEN D. KRASNER is Professor of Political Science at Stanford University and the editor of *International Organization*. His most recent monograph is *Asymmetries in Japanese-American Trade* (1987).

JOHN P. LEWIS is Professor of Economics and International Affairs at Princeton University. He has been working in and writing about India intermittently since 1959. He edited an Overseas Development Council volume, *Strengthening the Poor* (1988), and is currently finishing a collection of essays on Indian political economy and collaborating on a major history of the World Bank.

SUBRATA KUMAR MITRA is Senior Lecturer in Politics at the University of Hull, U.K. He is the author of *Governmental Instability in Indian States* (1978) and *Power, Protest and Participation: Local Elites and Development in India* (forthcoming). He has edited *The Politics of Positive Discrimination: A Cross-national Perspective* (1990) and *The Post-colonial State in Asia: The Dialectics of Politics and Culture* (1990).

PAUL D. HUTCHCROFT is a Ph.D. candidate in Political Science at Yale University. He is currently writing his doctoral dissertation, a study of relations between the state and dominant economic interests within the private commercial banking sector in the Philippines.

SAUL NEWMAN is Assistant Professor of Political Science at The American University, Washington, D.C. His current research concerns the influence of economic policies on electoral realignments and dealignments.

WORLD POLITICS

Vol. 43 • July 1991 • No. 4

CONTENTS

THE CONTRIBUTORS

Thomas Risse-Kappen is Assistant Professor of Government at Cornell University, on leave during 1990–91 at International Security Programs, Yale University. The author of *The Zero Option: INF, West Germany, and Arms Control* (1988), he is currently writing a book on intraalliance relations, particularly the European influence on U.S. foreign policy.

Peter Swenson is Assistant Professor of Political Science at the University of Pennsylvania and the author of *Fair Shares: Unions, Pay, and Politics in Sweden and West Germany* (1989). He is currently working on a comparative study of employers and politics in Sweden and the United States.

Cheryl Schonhardt-Bailey recently completed her Ph.D. in Political Science at the University of California, Los Angeles. She has been at the London School of Economics and Political Science and has been conducting a research project sponsored by the Economic and Social Research Council. In the fall, she will take up a position as Lecturer in International Relations at the University of Keele, Great Britain.

Forrest D. Colburn is Assistant Professor of Politics at Princeton University. He is the editor of *Everyday Forms of Peasant Resistance* (1989) and the author of *Managing the Commanding Heights: Nicaragua's State Enterprises* (1990). In 1988 he was a Fulbright Lecturer at Addis Ababa University.

Theodore Rueter has taught Political Science at Middlebury College, Georgetown University, and Smith College. He is coauthor of *Carter vs. Ford: The Counterfeit Debates of 1976* (1980) and *Teaching Assistant Strategies: An Introduction to College Teaching* (1990). His forthcoming books include an analysis of the Strategic Defense Initiative and an anthology of articles on international economic policy.

Thomas Kalil is an international economic analyst with the Washington law firm of Dewey Ballantine. He attended the Fletcher School of Law and Diplomacy at Tufts University, where he was an assistant editor of *Fletcher Forum*. He also worked as an issues analyst for the Dukakis presidential campaign.

Peter H. Smith is Professor of Political Science and Director of the Center for Iberian and Latin American Studies, and Simón Bolívar Professor of Latin American Studies at the University of California, San Diego. Prominent among his publications are *Argentina and the Failure of Democracy: Conflict among Political Elites, 1904–1955* (1974) and *Labyrinths of Power: Political Recruitment in Twentieth-Century Mexico* (1979). He is also coeditor of a five volume series entitled *Dimensions of United States-Mexican Relations* (1989–90).

THE POLITICAL ECONOMY OF TRANSITION

Japanese Foreign Direct Investments in the 1980s

By YOUNG-KWAN YOON*

INTRODUCTION

AN important issue in the area of international political economy since the 1970s has been explaining the success of the Japanese economy. Analysts have sought to identify causes of Japan's rapid economic growth in the post–World War II period in studies that may be assigned to two broad categories: the market and the nonmarket approaches.

According to the first approach, which concentrates on the concept of the market in explaining Japan's economic performance, the rapid economic growth of Japan was simply the result of the operation of market rationality and was not related to the uniqueness of the Japanese polity, society, or culture or to any other nonmarket factors. Hugh Patrick, Philip Trezise, and most other economists are proponents of this approach.[1] They assume that the concept of the market can be applied universally and that Japan is no exception.

The second, nonmarket approach focuses on concepts or aspects other than the market, such as Japan's state, institutions, psychology, culture, free-rider position, and so forth. Chalmers Johnson, for instance, formulated the concept of the "developmental state" to explain Japanese industrial development.[2] Though scholars working in this genre empha-

* This is a substantially revised version of a paper presented at the American Political Science Association Convention in Chicago in 1987. For their helpful comments and suggestions on earlier drafts of this paper, I would like to thank Charles F. Doran, Gary G. Hamilton, Jeffrey A. Hart, Robert Jackman, Bruce W. Jentleson, Edward J. Lincoln, George R. Packard, and Donald Rothchild. I remain entirely responsible, however, for the argument. I am also grateful for the research support of the University of California and to friends in Japan for their help during my research trip in Tokyo.

[1] Patrick, "The Future of the Japanese Economy: Output and Labor Productivity," *Journal of Japanese Studies* 3 (Summer 1977), 219–49; Trezise, "Politics, Government, and Economic Growth in Japan," in Hugh Patrick and Henry Rosovsky, eds., *Asia's New Giant: How the Japanese Economy Works* (Washington, DC: Brookings, 1976).

[2] Johnson, *MITI and the Japanese Miracle: The Growth of Industrial Policy, 1925–1975* (Stanford, CA: Stanford University Press, 1982). For an excellent review of various explanations, see chap. 1.

World Politics 43 (October 1990), 1–27

size different factors, all tend to stress the significance of nonmarket factors.

Researchers following both approaches tend, in explaining Japan's economic performance and comparing it with that of other advanced countries, to assume implicitly that characteristics of a society are more or less fixed and do not change significantly. This static assumption is more evident in the nonmarket approach than in the market approach, and it may help isolate the factors that distinguish the Japanese political economy from those of other advanced industrial countries at any given time. But the assumption of stasis leads analysts to overlook the fact that there have been rapid changes in various aspects of the Japanese political economy, especially after the early 1980s. Thus, it is important to analyze the nature of the change from a longitudinal perspective and to consider its implications for the dynamics of international political economy.

The basic question to be addressed is whether various aspects of the Japanese political economy are losing their particularity and coming to resemble the political economy of Western industrial societies. This question applies to such areas as government-industry relations, business-labor relations, finance-industry relations, intercorporate relations, and consumption and saving behavior, among others.[3] This paper attempts to answer the same question regarding Japanese foreign direct investment (FDI). I have analyzed the change in the nature of Japanese FDI and sought to clarify its causes and its implications for both the Japanese and the other world political economies. There is also an attempt to evaluate the impact of this change upon the Japanese economy.

Anticipating the conclusion, I argue that the pattern of Japanese FDI converged toward the Western style of FDI during the 1980s. Microeconomic explanations based on the concept of market failure have become

[3] In my opinion, particularlity seemed to prevail over generality in the earlier stage of Japanese economic development. However, as the Japanese accumulated wealth and entered the post-catch-up stage, generality began to appear more explicitly than before. For example the power and size of industry have grown so immense that it has become more difficult for the Japanese government to influence industrial behavior. See "Mighty MITI Loses Its Grip," *New York Times*, July 9, 1989, p. F1, and Daniel Okimoto, *Between MITI and the Market: Japanese Industrial Policy for High Technology* (Stanford, CA: Stanford University Press, 1989). Government-industry relations were affected by the gradually declining power base of the Liberal Democratic Party and the rising power of pressure groups. For a view which contends that Japan's political system has changed its nature and become more pluralist with labor's participation, see Ikuo Kume, "Changing Relations among the Government, Labor, and Business in Japan after the Oil Crisis," *International Organization* 42 (Autumn 1989), 659–88. In business-labor relations, the lifetime employment system began to erode and companies found it more difficult than before to seek sacrifices from their younger employees. See "Will Japan Really Change?" *Business Week*, May 12, 1986, p. 46. In finance-industry relations, debt financing has become less important, and big businesses depend more on foreign and domestic equity financing in the 1980s than they once did.

more pertinent in explaining Japanese FDI than were the macroeconomic explanations of the 1970s. This change reflects a fundamental shift of the Japanese economy from a trade-oriented to a foreign investment–oriented economy, the same pattern followed by Great Britain in the late nineteenth century and by the United States after World War II.

Understandably, most Western research on the Japanese political economy has paid little attention to Japanese FDI because the size of Japanese FDI was not important to Japan's economy until the late 1970s. It was after the rapid economic growth in the 1960s and 1970s and the subsequent accumulation of trade surpluses that the Japanese began to accelerate their FDI. Along with the change in its importance to the economy as a whole came a qualitative change in the nature of FDI.[4] This paper offers an explanation for these changes in attempting to fill the current gap in research on Japanese FDI.

As I have elaborated elsewhere, excessive foreign investments and the subsequent decline of productivity growth rate have been two important aspects of the behavior of global hegemonic powers since the nineteenth century.[5] It is evident that Japan cannot be called a hegemonic power in terms of capability and willingness.[6] Nevertheless, as its relative capability has increased, the importance of Japan has increased as well; and it will continue to do so in the world political economy. Thus, I compare the current Japanese experience in foreign investment with two classic examples—Great Britain before World War I and the United States after World War II. The latter section evaluates the impact of FDI on productivity growth.

THE CHANGE IN THE PATTERN OF JAPANESE FDI

The history of Japanese FDI can be divided into two phases according to the change in the size and nature of the investments. The first phase runs from roughly the latter half of the 1960s to the end of the 1970s. In

[4] FDI can be defined as a capital movement involving continuing control by the investor. See Charles P. Kindleberger, *American Business Abroad: Six Lectures on Direct Investment* (New Haven: Yale University Press, 1969), 3. I have not included foreign portfolio investments as a subject of discussion here, with the exception of the issue of *zai-tech* (sophisticated money management). This is because *zai-tech*, a type of portfolio investment, was made by industrial firms.

[5] Yoon, "Foreign Investment and Productivity: The British and American Cases" (Paper presented at the 1986 Annual Meeting of the American Political Science Association, Washington, DC, August 28–31, 1986).

[6] Keohane and Nye defined a hegemonic state as a state "powerful enough to maintain the essential rules governing interstate relations, and willing to do so." Robert O. Keohane and Joseph S. Nye, *Power and Interdependence: World Politics in Transition* (Boston: Little, Brown, 1977), 44.

this period Japan adopted an export-oriented strategy, and FDI took only a supplementary position in the Japanese economy. In the second phase, from the early 1980s to the present, not only the relative size, but also the nature, of FDI changed.[7] FDI assumed a much more important role in the Japanese economy, which shifted its emphasis from trade to investment. In this section, I review the development of this change in Japanese FDI and seek to identify its causes and assess its implications for the world political economy.

After the mid-1960s Japan began to have current account surpluses. As a result, gold and foreign exchange reserves increased from $2.0 billion in December 1964 to $15.2 billion in December 1971.[8] The Japanese government began to relax its control on outgoing FDI, which led to a noticeable increase in the amount of FDI from Japan. Except for the periods between 1968 and 1969 and between 1970 and 1971, the *amount* of FDI almost doubled or more than doubled every year in the period from 1967 to 1973.[9]

However, compared with U.S. FDI, the amount and relative importance of Japan's FDI in its economy was very small in the first phase. Whereas the cumulative total of U.S. FDI was $124 billion in 1975, Japanese FDI outstanding in all industries was only about $8.2 billion, which was equivalent to 1.7 percent of the gross national product, as shown in Table 1.[10]

Japanese FDI in the first phase was directed to three main areas, one of which was development of natural resources, mainly in developing countries. After the oil shock of 1973 Japan began to shift its emphasis from the strategy of importing vital resources to the strategy of development and import. Thus, Japanese corporations, with the active encouragement of the Japanese government, expanded investments in the Middle East and Asia. Increasing foreign reserve inflow enabled the Japanese to adop this strategy.

The second component of Japanese FDI in the first phase consisted c manufacturing investments, especially in low-technology sectors of de veloping countries in areas such as Asia and Latin America. For exan ple, as shown in Table 2, the total amount of manufacturing investmen

[7] The Japanese Economic Planning Agency also considered the years 1980–1981 as t beginning of another stage in the history of Japanese FDI. See Economic Planning Agen *Nihon Keizai no Genjō: Dorudaka Shūseika no Nihon Keizai* [The current situation of t Japanese economy: The Japanese economy adjusted for the high dollar] (Tokyo: Ōkura-s Insatsukyoku, 1986), 199–203.

[8] Bank of Japan, *Kokusai Shūshi Tōkei Geppō* [Balance of payments statistics monthly], J 1965, p. 117, and September 1972, p. 120.

[9] Ibid., August 1976, p. 47.

[10] The accumulated FDI from the United States in 1975 is from the U.S. Departmen Commerce, *Survey of Current Business*, February 1981, pp. 50–51, Table 7.

TABLE 1
JAPANESE FDI: PERCENTAGE OF GNP AND AMOUNT[a]

	1965	1970	1975	1980	1985	1988
Percentage	0.5	0.7	1.7	1.8	3.4	4.0
Amount[b] (billions of U.S. dollars)	0.4	1.5	8.2	19.5	45.0	113.1

[a] Two kinds of statistics are available for calculating Japanese FDI: one from the Ministry of Finance is based on the notification/approval, and the other is based on statistics for balance of payments. In this table, the accumulated FDI from Japan was calculated from Bank of Japan, *Kokusai Shūshi Tōkei Geppō*, June 1969, p. 47; August 1976, p. 47; September 1986, p. 53; December 1987, p. 61; and January 1989, p. 63. Calculation of GNP in dollars is based on International Monetary Fund, *International Financial Statistics*, July 1972, pp. 210–13; January 1978, pp. 204–7; August 1982, pp. 240–43; June 1986, pp. 286–89; and July 1989, pp. 314–16. Except for 1965, application of the exchange rate is based on the average annual par rate/market rate.

[b] Accumulated amounts based on balance of payments.

in developing countries was more than twice that in developed countries in the period of between 1972 and 1980.

The third component was investments in developed countries such as the United States and the European countries, mainly in the areas of commerce, banking, and services. The United States had an especially close trade relationship with Japan and initially imported labor-intensive goods such as textiles, toys, and sundries, and later, high-value goods such as televisions and automobiles. Thus, the Japanese needed to estab-

TABLE 2
COMPONENTS OF JAPANESE FDI[a]
(PERCENT)

	Contents	1951–71[b]	1972–80[b]	1981–84[b]	1988
[illegible]loped	Resources	10.3	7.0	2.5	2.9
	Manufacturing	9.2	11.4	15.0	16.4
	Other	31.3	26.9	32.6	42.2
[illegible]loping	Resources	14.3	14.4	9.9	5.5
	Manufacturing	18.2	24.0	12.2	10.4
	Other	16.5	16.2	27.9	22.6
[illegible]unt ([illegible]lions of U.S. dollars)		4.4	32.1	34.9	186.4

[illegible] column for cumulative amount in 1988 is calculated from MITI, *News from MITI* NR-366 (89-[illegible]89), 10. Investments in resources in this column include those in agriculture and forestry, [illegible] marine, and mining. All the other columns are from Economic Planning Agency, *Sekai Keizai [illegible]okuteki Seichō eno Kokusai Kyōcho o Motomete* [White paper on the world economy: In search [illegible]onal cooperation toward a continuous growth] (Tokyo: Ōkura-shō Insatsukyoku, 1985), 191.

[illegible] each period are based on the cumulative approval.

[illegible] rounding, percentages do not add up to 100.

lish market networks to take care of inventories, advertising, and after-sales services.[11]

When we narrow our focus to Japanese FDI in the manufacturing sector, we find that it contrasts in several ways with that of other advanced industrial countries.[12] First, as already noted, investments in developing countries were much greater than those in developed countries. Investments in developing countries were mainly in Asia and Latin America, in low-technology, labor-intensive (and declining) industries such as textiles, sundries, metal products, electrical machinery, and chemicals.

Second, the Japanese companies that invested abroad were not necessarily large oligopolistic firms. In this respect Japanese firms differed from firms based in the United States and the countries of Western Europe. According to a study by the Ministry of International Trade and Industry (MITI) in March 1975, sixty-two large firms and forty-five small or medium-size firms engaged in FDI in the textile industry.[13] Almost half of the total number of Japanese FDIS were made by small and medium-size enterprises (those with either a total number of employees of three hundred or fewer or paid-in capital of 100 million yen or less) in 1975.[14]

Third, general trading companies participated actively in FDI, even in manufacturing industries. Ozawa argued that most of the Japanese overseas manufacturing ventures were, by Western standards, still immature in their scale of operation, technological sophistication, and financial capability. General trading companies helped to overcome these handicaps.[15] When there was a potential threat of loss in the export market, it was the trading companies that first brought it to the attention of the manufacturers and helped them begin production abroad. The manufacturing firms looked to the trading companies for knowledge and expertise in international business, financial resources, and local know-how and in return they usually permitted an equity position for trading companies.[16] Thus, in the first phase a unique industrial network among Japanese firms had an important influence on their FDI behavior.[17]

[11] Terutomo Ozawa, *Multinationalism, Japanese Style* (Princeton: Princeton Univers Press, 1979), 22.

[12] For a good analysis of Japanese FDI in each region by the mid-1970s, see Sueo Sekiguc *Japanese Direct Foreign Investment* (Montclair, NJ: Allanheld, Osmun, 1979), esp. 53–58.

[13] See Sekiguchi (fn. 12), 57, Table 3-4, based on MITI, *Overseas Activities of Japanese I terprises*, 1976.

[14] Ozawa (fn. 11), 26.

[15] Ibid., 25, 30.

[16] M. Y. Yoshino, "Japanese Foreign Direct Investment," in Isaiah Frank, ed., *The Japa Economy in International Perspective* (Baltimore, MD: Johns Hopkins University Press, 19 262–63.

[17] See Lawrence B. Krause and Sueo Sekiguchi, "Japan and the World Economy, Patrick and Rosovsky (fn. 1), 392–93, and Sekiguchi (fn. 12), 56–58, 139–43.

Considering the above factors, it is evident that Japanese FDI at this time differed in several ways from that of the other advanced industrial countries. A few Japanese scholars tried to explain this phenomenon from a theoretical perspective. Kojima and Ozawa, for example, emphasized the particularity of Japanese FDI. They approached the phenomenon from a macroeconomic perspective and explained its particular characteristics in relation to the change in the macroeconomic necessity of the Japanese economy.

In contrasting Japanese FDI with U.S. FDI, Kojima found that Japanese investments aimed mostly at exploiting natural resources in resource-rich countries and low wages in labor-abundant countries. These investments usually led to the increase of exports to Japan or to a third market from host countries. According to Kojima, Japanese FDI supplemented trade and followed the principle of comparative advantage from the macroeconomic perspective. However, U.S. manufacturers invested in highly sophisticated, technology-based products to increase their sales to the market of the host country. This style of investment tended to replace exports from the United States and ran counter to the principle of comparative advantage. Kojima called the Japanese style of FDI a trade-oriented investment and the U.S. style of FDI an anti-trade-oriented investment.[18]

Ozawa argued differently, that these characteristics of Japanese FDI were associated with Japan's industrial transformation from labor-intensive, low-tech, and low-growth sectors to high-tech and high-growth sectors. In the process of structural adjustment, such macroeconomic forces as factor scarcities at home, an uncertain foreign supply of key resources, and a decline in the competitiveness of Japanese labor-intensive goods prompted Japanese firms to invest abroad. Thus, Japanese FDI differed from U.S. FDI, which had been propelled by the growth of the internal capacity of each individual firm.[19]

Kojima, who approached the phenomenon of FDI from a macroeconomic perspective, assumed the rationality of Japanese investment behavior and justified it from a perspective of world welfare. Ozawa also assumed the rationality of Japanese investment behavior in concentrating on low-tech and labor-intensive industries. According to him, technology released from the contracting low-tech sector would be transferred to the less-developed host countries and would increase world welfare.[20]

[18] Kiyoshi Kojima, *Direct Foreign Investment: A Japanese Model of Multinational Business Operations* (New York: Praeger, 1978), esp. chaps. 4, 5, and 6.

[19] Ozawa (fn. 11), 70.

[20] Ibid., 63. However, Japanese investors were often criticized for their reluctance to transfer even lower-grade technologies to host countries because of their fear of a boomerang

These two explanations emphasized that Japan's FDI was the result of rational behavior on the part of Japanese industry in following macroeconomic necessities; thus, it was essentially unique and different from Western FDI.

Western scholars, however, tended to overlook these differences between Japanese and U.S. FDI. They assumed that theories of FDI could be applied universally to explain even Japanese FDI. The mainstream theoretical explanations in the United States were based on the microeconomic approach, which viewed FDI as resulting from market failure. For instance, Stephen Hymer, Charles Kindleberger, and Richard Caves pointed out in their respective publications that FDI usually occurred in the oligopolistic market. They argued that oligopolistic firms had some advantages over local corporations in that they could compensate for transaction costs or overcome barriers to entry into the host markets.[21] Even Raymond Vernon's product-cycle theory turned our attention to the development of new technology and products at the initial stage of the cycle.[22] Usually large oligopolistic firms generated new technologies and products after all. In this sense, the product-cycle theory also shared the assumption of the market-failure approach.

These theories reflected the characteristics of American multinationals in the 1960s and 1970s, when the United States held the majority of world FDI. American scholars either assumed that these theories could explain Japanese FDI or did not show much interest in this issue because of the small size of Japanese FDI in the 1960s and the 1970s. John E. Roemer, for example, argued that the trade-oriented nature of Japanese investments was just a fortuitous circumstance and that as the Japanese economy matured, the corporate behavior of Japanese firms and market patterns would come to resemble the Western style of FDI.[23] He argued that

---

effect. In addition, Japanese investments in low-tech areas were often accompanied by th export of pollution.

[21] Hymer, *The International Operations of National Firms: A Study of Direct Investmen* (Cambridge: M.I.T. Press, 1976); Caves, "International Corporations: The Industrial Eco nomics of Foreign Investment," *Economica* 38 (February 1971), 1–27. Kindleberger empha sized the importance of this approach and claimed that most of the other recent explanatior were derived from Hymer's thesis. Charles P. Kindleberger, *Multinational Excursions* (Can bridge: M.I.T. Press, 1984), 180.

[22] Vernon, "International Investment and International Trade in the Product Cycle *Quarterly Journal of Economics* 80 (1966), 190–207.

[23] Roemer, *U.S.-Japanese Competition in International Markets: A Study of the Trade-Inve ment Cycle in Modern Capitalism*, Research Series No. 22 (Berkeley: Institute of Internatior Studies, University of California, 1975), 206–7, 210. In explaining Japan's international co petitiveness, he suggested a four-stage hypothesis, which, I believe, has a high degree explanatory power from the long-term perspective. For this hypothesis, see pp. 131–41. U fortunately, Roemer did not explain the mechanism from a theoretical perspective.

> although the data . . . and the observations about Japanese manufacturing investment drawn from them are consistent with Kojima's hypothesis, these data and observations can be explained by certain fortuitous circumstances which have nothing to do with a more "rational" policy of Japanese investment. Furthermore, it is probable that the pattern will change and that Japan, too, will invest strongly at the "top of the product cycle." The fortuitous circumstances are: (1) Southeast Asia has been Japan's main "sphere of influence," and labor-intensive investment is the natural type of investment to make in that area; (2) the existence of trading companies has facilitated early, small-scale investments in non-competitive industries; (3) a large, unsaturated domestic market made foreign expansion in the advanced sectors less critical during the 1960s. Thus Japanese foreign investment in the 1960s was atypical.[24]

Interestingly enough, the new trends in Japanese FDI in the second phase, after the early 1980s, bore out Roemer's prediction. The particular characteristics of Japanese FDI in the first phase were overwhelmed by the newly emerging Western style of Japanese FDI. What accounts for the change that occurred in the second phase of Japanese FDI? An important factor was that the conservative attitude of the Japanese people toward FDI underwent a rapid change. For example, the Japanese government publicly announced that the coming years would be a period of investment instead of trade: "In a long-term view of the future of the economy, it is inevitable that Japan will take the route from being a major trading power to being a major power in direct investment."[25] Moreover, the Ministry of Finance strongly backed the loan program of the Export-Import Bank of Japan for Japanese investors going abroad.[26] Also, a range of tax incentives has been developed for potential investors. These changes altered the nature of Japanese FDI.

First, reflecting the fundamental changes in attitude and policies toward FDI, the amount of Japanese FDI began to increase rapidly. On the basis of approval, cumulative FDI has increased by five times, from some $36 billion in 1980 to $186.4 billion in 1988.[27] It is noteworthy that this increase occurred at a time when the United States had begun to decrease its share of the world FDI. If this trend continues, Japanese FDI will dis-

[24] Ibid., 176–77.

[25] Economic Planning Agency, *2000 Nen no Nihon: Sekai Keizai, Takyoku Antei o no Dohyō* [Japan in the year 2000: Guideposts to the world economy, multipolar stability] (Tokyo: Ōkura-shō, 1982), 69.

[26] Outstanding loans of this program have increased from 62.4 billion yen in 1977 to 258 billion yen in 1983. See "Japan '85: Getting Round Protectionism by the Direct Route," *Far Eastern Economic Review*, June 13, 1985, p. 83.

[27] MITI, *News from MITI* NR-366 (89-9) (July 1989), 2–3. As shown in Table 1, the cumulative total of Japanese FDI on the basis of balance of payments statistics is smaller and reached $113.1 billion by the end of 1988.

place that of the United Kingdom to account for the world's second-largest proportion of FDI, following that of the United States.

Second, the most important change was the shift from investments oriented to resources and cheap labor in developing countries to the market-oriented, technology-based investments in industrialized countries such as the United States and those in Western Europe. According to the 1989 edition of the *Economic White Paper*, Japanese FDI by about 1980 was directed mainly toward Asia, to resource-dependent industries and other industries in which Japan was at a comparative disadvantage. In recent years, however, most of Japanese FDI was made in industries of Japan's export competitiveness, such as electrical machinery, general machinery, and transportation equipment in industrialized countries.[28] For example, as shown in Table 3, the relative share of FDI in processing and assembling in North America increased from 8.5 percent in 1975 to 37.6 percent of the total manufacturing investments in 1985. The reverse has occurred with investments in basic industries in Asia; these decreased from 24.3 to 5.2 percent between 1975 and 1985.

This change was accompanied by the increasing importance of large oligopolistic firms in FDI.[29] All major automakers have set up manufacturing operations in the United States by now. Even Toyota, which had been reluctant to produce abroad, followed the lead of Nissan, Honda, and Mitsubishi and began production in the United States. In the steel industry, Nippon Kokan recently invested in and gained control of National Steel. In the high-tech electronics industry, many large companies, including NEC, Sony, and Fujitsu, are producing everything from personal computers to cellular mobile telephones in the United States.

The rapid increase in the share of FDI in machinery, transportation equipment, and electrical industries, as depicted in Table 3, reflects the rising importance of FDI made by high-tech, capital-abundant oligopolistic firms. Most of these investments were made in the United States. Japan's FDI in the U.S. manufacturing industry is expected to grow by 14.2 percent annually until the year 2000, which will lead to a tenfold increase over the level of the early 1980s.[30] Table 4 shows the trend to

[28] Economic Planning Agency, *Keizai Hakusho 1989: Heisei Keizai no Kadode to Nih[...] Keizai no Atarashii Chōryū* [Economic white paper 1989: The start of the Heisei economy a[...] the new trend of Japanese economy] (Tokyo: Ōkura-shō Insatsukyoku, 1989), 198–200.

[29] There are no available data for measuring the increasing importance of large firn[...] However, a rough measure like dollar amounts of FDI flow per case confirms this trend. [...] has increased from $2.5 million per case in 1976 to $9.6 million per case in 1988 in the electr[...] electronics industry, from 3.6 to 10.0 for transportation equipment, and from 2.5 to 7.7 [...] manufacturing as a whole. Calculated from MITI (fn. 27), 9, Table 3, and Keiji Nakata[...] "Trends in Japan's Direct Investment Abroad in FY 1986," *EXIM Review* 8 (March 19[...] 76–103, at 90, Table 7.

[30] "Japan in America," *Business Week*, July 14, 1986, p. 46.

TABLE 3
DISTRIBUTION OF JAPANESE FDI IN MANUFACTURING
(PERCENT)

	Region	1975	1980[a]	1985[a]
Processing and Assembling[b]	North America	8.5	12.7	37.6
	Machinery	3.1	1.8	10.3
	Electrical	4.7	9.8	17.1
	Transportation equipment	0.7	1.1	10.2
	Asia	9.3	7.7	11.9
	Other	14.0	14.1	13.9
		(31.8)	(34.4)[c]	(63.4)
Basic[b]	North America	22.3	9.9	10.9
	Asia	24.3	30.7	5.2
	Other	12.2	19.8	11.6
		(58.8)	(60.4)	(27.7)
Other	(All regions)	9.4	5.2	8.8

SOURCE: Economic Planning Agency, *Keizai Hakusho 1986: Kokusaiteki Chōwa o Mezasu Nippon Keizai* [Economic white paper 1986: Japanese economy aiming at international harmony] (Tokyo: Ōkura-shō Insatsukyoku, 1986), 199, 515.

[a] Due to rounding, percentages do not add up to 100.

[b] Processing and assembling industries include industrial machinery, electrical machinery, transportation equipment. Basic industries include food, textiles, wood and wooden products, pulp and paper products, chemicals, iron and steel, and nonferrous metals.

[c] Numbers in this column total 34.5; 34.4 is the total given in the source.

ward increased Japanese FDI in the industrialized West and its decline in developing countries, which confirms the pattern of Japanese FDI becoming more and more oriented to the export market.[31] Thus, Japanese FDI has increasingly come to resemble the Western style of FDI.

Another noticeable change has been the rapid increase in nonmanufacturing investments after 1985, especially in the United States. The most important factor contributing to this change was the increase of investments in finance and insurance, which constituted 40 percent of the annual outflow of nonmanufacturing FDI in 1985 and 1986 and 43 percent in 1987.[32] The flow of FDI in finance and insurance almost doubled annually between 1982 and 1986, and it continued to increase

[31] More than 90% of total sales of Japanese production in North America and Europe was [illegible] the local market in 1987. The ratio was as much as 60% in Southeast Asia. See Economic Planning Agency (fn. 28), 202, Table 3-2-10(1).

[32] Ministry of Finance, "Foreign Direct Investment by Industry," as quoted in Onishi Katsuaki, "Kaigai Shinshutsu no Kyuzo de Keizaiwa 'Kūdoka' Shinaika" [No "hollowing-out" of the economy in spite of the rapid foreign expansion?], *Keizai* (May 1989), 48–57, at 49, and Nakatani (fn. 29), 90.

TABLE 4
REGIONAL DISTRIBUTION OF JAPANESE FDI[a]
(PERCENT)

	North America	*Europe*	*Oceania*	*Latin & S. America*	*Asia*	*Middle East*	*Africa*
1975	27.7	10.1	5.5	11.3	33.5	6.1	5.8
1980	34.0	12.3	9.5	12.5	25.3	3.4	3.0
1985	45.0	15.8	4.3	21.4	11.7	0.4	1.4
1987	46.0	19.7	4.2	14.4	14.6	0.2	0.8

SOURCE: Arranged from Ministry of International Trade and Industry (MITI), *Keizai Kyōryoku no Genjō to Mondaiten, 1985* [The current situation of economic cooperation and problem areas, 1985], March 1986, p. 245; Japan External Trade Organization (JETRO), *Sekai to Nihon no Kaigai Chokusetsu Tōshi* [The world's and Japan's foreign direct investments], 1987, pp. 376–77; 1988, pp. 384–85; 1989, pp. 393–94.
[a] On the basis of approval.

rapidly afterward.[33] This trend reflects the fact that Japanese financial institutions have been following the rapid expansion of Japanese manufacturing firms abroad, which tended to depend exclusively on foreign branches of Japanese banks for financial services.[34] Though the increase in nonmanufacturing investments was a secondary phenomenon caused by the expansion of manufacturing investments, it reveals the increasing internationalization of Japanese finance as well as industry.

The above-mentioned changes were accompanied by a decline in the role of trading companies. Initially, as the foreign market expanded, Japanese manufacturing companies needed to have direct control in pricing and marketing in order to provide the kind of sales engineering and service required for the distribution of sophisticated products.[35] In addition, as a number of host countries began to industrialize and achieve import substitution, Japanese firms were pushed to expand their local production process from limited, small-scale operations, such as assembly, to upstream or downstream operations.[36] Some manufacturing firms also acquired experience in international business and could make greater financial, technical, and managerial commitments. Thus, the rol

[33] Nakatani (fn. 29), 90.
[34] See the survey " 'Sihon Yushutsu Koku' ka, Ginkō no Kokusaika oyobi Kyatchi Appu [Becoming a capital-exporting country: Internationalization of banks and catch-up], *Keiz Kenkyū* 37 (July 1986), 258–75, at 271, Table 18. Data show that there is a consistency in t distribution between manufacturing investments and investments in finance and insuran See p. 272, Table 19.
[35] James C. Abegglen and George Stalk, Jr., *Kaisha, the Japanese Corporation* (New Yo Basic Books, 1985), 246.
[36] A good example is the synthetic fiber industry. See Yoshino (fn. 16), 267. Yoshino arg that this trend began around 1975.

of trading companies began to weaken, and manufacturing firms began to take the initiative in foreign production.[37] In this way, Japanese manufacturers began to make direct marketing investments themselves and relegated general trading companies to the low-margin commodity area.[38] This is another aspect of vanishing particularity—a weakening of the influence of the intercorporate network on the behavior of Japanese FDI.

All these changes indicate the increasingly Western style of Japanese FDI. What is the cause of this change? First, newly industrializing countries were changing. As they developed their own labor-intensive and low-technology industries—on which Japanese FDI had traditionally focused—the relative importance and profitability of these industries became less significant for Japanese investors, who were thus forced to move to the upscale and/or high-technology areas. Even in the automobile industry, the newly industrializing countries began to occupy the lower end of the market. For this reason, Japanese automobile manufacturers have had to enter into joint ventures with these new challengers or move into upper-market streams and leave their share of the lower-end market to companies such as Hyundai.[39]

Likewise, Japanese steel producers, who once devastated the U.S. steel industry with low-priced exports, have been hard-pressed to lower prices or risk losing business to Brazil or Korea.[40] Similar trends are expected to occur in the semiconductor industry. Moreover, new Japanese FDI in resource development has been steadily declining because of the shift in emphasis from resource-consuming industries to high-tech industries and because of declining domestic economic growth in Japan.[41]

Second, in these high-technology industries, Japanese exporters found themselves facing increasingly severe trade restrictions from industrialized countries. To circumvent these trade barriers, Japanese firms had no alternative but to produce directly in the foreign markets. For example, after the United States imposed an import quota on consumer electronics goods such as color television sets, all major Japanese producers moved to set up U.S. assembly operations. Similarly in the automobile

[37] Ibid., 266–68.

[38] Abegglen and Stalk (fn. 35), 246.

[39] Sato Sadayuki, "Nihon Keizai mo 'Kūdoka' no Osore: Shindankai Mukaeru Nihon Kigyo no Takokuka" [The fear of the hollowing-out of the Japanese economy: Toward a new stage of multinationalization of Japanese business], *Ekonomisuto*, February 4, 1986, p. 18. See also "Japan Picks Winners to Hurdle the Fences," *Far Eastern Economic Review*, September 4, 1986, p. 61.

[40] "Will Japan Really Change?" *Business Week*, May 12, 1986, p. 49.

[41] Kiyoshi Kojima, "Japanese-Style Direct Foreign Investment," *Japanese Economic Studies* 14 (Spring 1986), 52–82, at 56.

There thus arose an imbalance between the rigidity of Japanese institutions and a rapidly changing new reality in the distribution of wealth in the world political economy. The Japanese have not been ready and willing to adjust their life-style to the new situation. Rather, they have clung to the mercantilist way of life that made them a successful challenger.[48] Unfortunately, in an interdependent world, the imbalance could not remain as a purely Japanese problem. It was bound to have an impact abroad later because it created problems in the flow of goods and capital in the international economic system. In the early 1970s the Japanese economy began to be regarded as a destabilizing factor in the world economy, and other economic forces started to bring pressure upon Japan to adjust itself to the new world economy. By way of response, the Japanese government found no alternative but to encourage foreign direct and portfolio investments and to give back what Japan earned through its exports. The result has been a huge increase in the amount of foreign investments in the 1980s, a trend that will continue in the near future.[49] In other words, the very success of the trade-oriented strategy brought about its own demise and the shift to the investment-oriented strategy.

What are the implications of being a nonhegemonic power in the overall process of transition from a trade-oriented economy to an investment-oriented economy? As a nonhegemonic power, Japan has had a target: the United States. Having a target on which to focus, the Japanese staved off complacency and inefficiency in the operation of institutional mechanisms for a while; instead, they enjoyed the advantage of the "economic backwardness" of latecomers.[50] But the very fact that Japan was a nonhegemonic power limited its range of strategic choices. Similar changes from a trade-oriented economy to an investment-oriented economy in the British and American cases were caused by the development of complacency and the inefficiency in institutional mechanisms. In con-

[48] The same analogy can be applied to the case of the United States. There is an imbalance between the American mind and institutions and the country's wealth. Caught by the inertia of the past, the American government and people keep consuming more than they can afford.

[49] There is no clear evidence that the recent boom in domestic demands in 1988 and 1989 was caused by institutional reforms. Rather, the boom seems to be a temporary phenomenon resulting from the appreciation of the yen and cyclical factors. See "The Consumption Boom in Japan," *Tokai Monthly Economic Letter* 125 (June 1989), as quoted in *Japan Update* (Autumn 1989), 9–14.

[50] I use this term keeping Gerschenkron's usage in mind. Such distinctive characteristics of industrial latecomers as the speed of industrialization, productive and organizational structure of industry, and etatist ideologies can be understood as the result of creative response to or application of the past route of industrial revolution of the forerunner countries. See Alexander Gerschenkron, *Economic Backwardness in Historical Perspective* (Cambridge: Harvard University Press, 1962), 6–11, 22–29.

trast, this transition in Japan was initiated by external forces that would not tolerate the imbalance between the rigidity of Japanese institutions and the new reality of world political economy.

After the Change: The Consequence of Increasing FDI

As seen in the cases of Great Britain and the United States, foreign investments negatively affect growth in domestic productivity.[51] Industrial managers are able to capture profits by shifting production abroad in the short term, but the diversion of capital from what is traditionally the most productive export sector inevitably leads to an increase in unemployment, the shift of unemployed labor to lower-productivity growth sectors, slower or no accumulation of production experience and technology, less enthusiastic pursuit of innovation, and more financially oriented industrial management.[52] The mirror image of this mechanism is the expansion of the service sector and the phenomenon of deindustrialization, that is, the depression of productivity growth at the aggregate level. Contrary to conventional wisdom, the short-term profit of an individual firm is very different from the long-term productivity growth of the aggregate economy.

Is there any evidence to show that this mechanism is working and will continue to work in Japan? Although it is too early to analyze the full effect of FDI on productivity growth, evidence is already accumulating that supports the hypothesis. In this section, I focus on this evidence.

The era of FDI has just begun in Japan, which is still in the midst of transition from a trade-oriented to an investment-oriented economy. According to the Economic Planning Agency, the ratio of the overseas production of manufacturing industries to their total output, which was 3.4 percent in 1982 and around 5 percent in 1986, is expected to increase to 20 percent by the year 2000.[53] FDI is thus affecting the evolution of the

[51] For examples of parallel arguments, see J. M. Keynes, "Foreign Investment and National Advantage," *The Nation and the Athenaeum* 35 (August 9, 1924), 584–87; Marvin Frankel, "Home versus Foreign Investment: A Case against Capital Export," *Kyklos* 18 (March 1965), 411–33; Robert Gilpin, *U.S. Power and the Multinational Corporation* (New York: Basic Books, 1975); Sidney Pollard, "Capital Exports, 1870–1914: Harmful or Beneficial?" *Economic History Review* 38 (November 1985), 489–514; and William P. Kennedy, *Industrial Structure: Capital Markets and the Origins of British Economic Decline* (Cambridge: Cambridge University Press, 1987).

[52] The trend toward financially oriented management was encouraged even more by foreign portfolio investments since the 1980s. It is clear, however, that the procedure of FDI itself brings its own mechanism to bear on this trend, as we can see from the American experience of offshore production in electronics and other industries in the 1960s and 1970s. Evidence suggests that a similar mechanism is at work in the case of Japanese FDI.

[53] "Employment Opportunities Follow Production Bases out of Japan," *Japan Economic Journal*, July 19, 1986, p. 17.

entire Japanese economy, with the result that increasing FDI is displacing exports. A good example is the electronics industry. In the 1970s the export of Japanese color television sets increased rapidly, but, then, from July 1977 to June 1980 the United States imposed export restraints on Japan. In order to avoid these restraints, Japanese manufacturers of color televisions began to transfer their production to the United States. While Japanese production in the United States rose, exports from Japan began to decline, and in 1978 Japanese production in the United States caught up with exports. This negative relation between exports and foreign production continued until recently. Another typical example is the case of videotape recorder (VTR). The exports of VTRs to countries of the EC began to decline in the same year that Japanese manufacturers began production in those countries.[54]

In the machinery industry, the share of the foreign production to the total supply for foreign markets is expected to be about 33 percent in 1990, and this is expected to increase to 46 percent by 2000. The share of exports, according to the report by a Japanese machine industry organization, is expected to decrease by the same rate.[55]

The same thing is happening in the automobile industry. Japanese companies have also been pressured by host governments to raise the local content ratio in the production of automobiles. The negative relation between FDI and exports in the automobile and parts industries became more evident after 1986.[56] A private research institute estimated that the increase in production of motor vehicles by the Japanese companies in the United States would reduce Japan's 1992 trade surplus with the U.S. by $7.3 billion from 1988 levels.[57] Furthermore, there has been an increase in reverse imports from the foreign affiliates of Japanese multinationals to the home country. For example, in FY 1987 the total value of reverse imports was estimated at $9.4 billion, which accounted for 11.5 percent of Japan's total manufactured imports.[58]

The displacement of exports by FDI has led to an increase in unemployment in the traditionally most efficient industries. Although the boom in capital investment in 1989 may conceal the problem for the time being, the problem clearly exists, and it is of a structural nature in two respects.

[54] See the two graphs that show the negative relationship between local production and exports (color TV in the United States and VTR in the EC market) in MITI (fn. 42), 193.

[55] Japan External Trade Organization (JETRO), *Sekai to Nihon no Kaigai Chokusetsu Tōshi* [The world's and Japan's foreign direct investments], 1988, 41.

[56] Economic Planning Agency (fn. 28), 208–10, 578: see esp. Table 3-2-12.

[57] Fuji Research Institute Corporation, *Fuji Economic Review* (July–August 1989), 15.

[58] MITI, International Business Affairs Division, "The 18th Survey on Japanese Business Activity Abroad," April 27, 1989, p. 3.

First, if all the unemployed people from the FDI-intensive industries could be employed in the area of higher-productivity growth industries, we would not have to worry about the hollowing-out phenomenon. However, this does not appear to have happened. Rather, the available evidence shows that most unemployed people were absorbed into lower-productivity growth sectors. For instance, in the years from 1985 to 1987, when FDI was most intensive, the number of people employed in the manufacturing sector decreased by 280,000, while the number in the service sector increased by 810,000 (those in retail and restaurant businesses increased by 480,000).[59] The rates of productivity growth in services, distribution, agriculture, and the construction sectors are usually lower than those in the manufacturing industries. The shift of workers from a manufacturing to a service sector or from a higher-to a lower-growth manufacturing sector will depress the productivity growth of Japan at the aggregate level.

Second, the negative effect of the hollowing-out of export industries on productivity growth will be greater in Japan than it was in the United States in recent decades. Compared with the pattern of economic development in the United States, Japanese development has been more unbalanced and has depended more on the development of a few particular export industries. For instance, even though U.S. manufacturing industries were hollowed out by FDI, the United States still leads in many sectors such as basic research, software, and space development. Japan, however, has been almost totally dependent on a few major manufacturing industries.[60] As a consequence, the hollowing-out of these export industries will more deeply affect productivity growth at the aggregate level in Japan than it did in the United States.

Furthermore, increasing FDI will negatively affect productivity growth through the destruction of the traditionally efficient domestic structure of production, an important source of Japanese competitiveness. A good example was the reluctance of Toyota to invest abroad. Toyota's efficient production system in Toyota City consisted of the "kanban, or just-in-time method, a good manufacturing location, amicable industrial relations and cooperative auto part suppliers."[61] FDI weakens these positive factors and relationships at home. This was why Toyota, Japan's largest automaker, was the last major foreign investor from the Japanese auto-

[59] Ministry of Finance, *Bureau of International Finance Yearbook 1988* (Tokyo: Ōkura-shō, 1988), 91.

[60] "Danger of Industrial Hollowing-Out," *Japan Economic Journal*, October 18, 1986, p. 6. See also "Producing Abroad Faces Problem," *Japan Economic Journal*, December 27, 1986, p. 6.

[61] "Toyota's Unique Method of Production Faces Real Test in North America," *Japan Economic Journal*, December 21, 1985, p. 18.

mobile industry. Now that the production base has been transferred, advantages of scale and reliable, cost-effective relationships with parts suppliers will no longer be available at home as they had been before.

Overseas production also resulted in fewer orders for parts suppliers. Thus, some parts manufacturers followed the main producers and moved abroad, but the number of these firms was small because most of them lacked the financial capability to make the move. According to a local financial institution, of the ninety subcontractors of the Hiroshima based Mazda Motor Corporation, only about ten companies could begin overseas production by early 1986.[62] Those suppliers that did not have enough financial power could not move out and had either to reorient their product to the remaining domestic market or to change their product line. This process inevitably leads to inefficiency and higher adjustment costs for the entire economy. In addition, the firing of workers in subcontractor companies and related industries results in the decline of company loyalty, deterioration in the quality of work of the remaining workers, and the weakening of "lifetime employment," one of the key factors of the "Japanese management style." The continuously increasing ratio of the outsourcing of parts and materials from offshore suppliers to the domestic supply will further aggravate this problem.[63]

Investment overseas is a drain on total Japanese investments, to the disadvantage of the development of the domestic economy.[64] In such FDI-intensive industries as the automobile and parts industries, only 8 percent of twenty-five companies that planned to invest abroad felt that domestic investments in plants and equipment were more important than foreign investments in plants and equipment. Twelve percent felt that foreign investments were a little more important than domestic investments, whereas 48 percent regarded them as just as important as domestic investments. Thirty-two percent perceived no significant difference between the two.[65] In addition, we can find a striking contrast in the patterns of investment between the earlier period of high growth with small FDI and the later period of low growth with large FDI. For instance, pri-

[62] These ten companies are known to have annual sales of 20–40 billion yen each. See "Exporters Attempt New Businesses, Takeovers, to Combat High Yen," *Japan Economic Journal*, May 24, 1986, p. 1.

[63] For the increasing ratio of outsourcing, see Economic Planning Agency (fn. 28), 215. It was reported that the unemployment rate in Japan had risen to 3% in January 1987, the highest level since monthly reporting began in 1953. See "Mounting Conflict over Trade Looms for U.S. and Japan," *New York Times*, March 9, 1987, p. D4.

[64] Abegglen and Stalk (fn. 35), 257.

[65] Economic Planning Agency, Bureau of Analysis, *Kokusaiteki Senryaku o Susumeru Kigyō Kōdō* [Business behavior pursuing international strategy] (Tokyo: Keizai Kikaku-cho, 1987), 281–82.

vate investments in plants and equipment increased by an average of 14.4 percent annually during the time of high growth between 1952 and 1973, but the annual growth slowed down to 3.3 percent between 1974 and 1987 (or 7.3 percent between 1980 and 1987).[66] The capital investment boom in 1989 is comforting for the Japanese, but it may not last long.[67] Also the ratio of gross private capital formation to GNP decreased from 35 percent in 1970 to 32 percent in 1980 and 28 percent in 1985, while the ratio of the (accumulated) FDI to GNP increased from 0.7 percent in 1970 to 1.8 percent in 1980 and 4.0 percent in 1988.[68]

The more fundamental reason for a negative effect of FDI on productivity growth is that FDI will lead to the loss of production experience. Accumulation of production experience has been the core of Japan's long-term industrial strategy, which has emphasized utilizing the production experience and technology of lower-level and lower-value-added products to develop higher-level and higher-value-added products. This issue needs to be examined in more detail.

When people, including many corporate managers, worried about the hollowing-out effect, a typical response from the other end of the spectrum was the division-of-labor argument. For instance, about four-fifths of Japan's sixty-three major corporate heads said that there would be a hollowing-out of the manufacturing industry as a result of the shift to overseas production. They expected this hollowing-out to occur especially in the automobile and electronics industries, the two most important industries in Japan.[69] However, a Japanese industrial manager's comment is reminiscent of the arguments made in the 1960s and 1970s by some American managers and economists who justified FDI as a way of doing "brain work" at home and production abroad: "Japan will certainly see a hollowing out in production. But it is necessary to carry out research and development, develop prototype products and establish mass-production know-how in Japan."[70]

[66] The annual growth rate for the period 1952–1973 is from Hugh Patrick and Henry Rosovsky, "Japan's Economic Performance: An Overview," in Patrick and Rosovsky (fn. 1), 17. The rate between 1974 and 1987 was calculated from Economic Planning Agency, *Keizai Hakusho 1988: Naijugata Seichō no Jizoku to Kokusai Shakai e no Kōken* [Economic white paper 1988: Continuing the domestic demand-oriented growth and contributing to the international society] (Tokyo: Ōkura-shō Insatsukyoku, 1988), 440.

[67] See *Economist*, February 10, 1990, p. 63; *Wall Street Journal*, February 15, 1990, p. 1.

[68] The calculation is based on International Monetary Fund, *International Financial Statistics* (July 1972), 210–13; (January 1978), 204–7; (August 1982), 240–43; (June 1986), 286–89. (Application of the exchange rate is based on the average annual par rate/market rate, except for 1965.) See Table 1.

[69] "Nihon Keizai Shimbun Survey: Corporate Heads See Hollowing-Out of Industry Due to Shift Overseas," *Japan Economic Journal*, June 7, 1986, p. 1.

[70] See "Hitachi Chemical to Keep Research at Home, Shift Production Overseas," *Japan Economic Journal*, October 18, 1986, p. 21.

Nevertheless, a more important factor than brain work is the produ tion experience that raises the overall level of skills in a population an leads to productivity growth.[71] *Japanese producers in the 1960s and 1970 could develop the capacity to take advantage of new and innovative idea in making competitive goods by keeping production experience at home* However, with increasing FDI, the Japanese will voluntarily export or eventually be forced to export the production experience, and their con*centration on research and development (R & D) and the development of new technologies will thus not automatically lead to the efficient pro*duction of competitive goods. This was the case of FDI-intensive industries like the consumer electronics and machine-tool industries of the United States in the 1960s and 1970s.

The 1989 edition of the *Economic White Paper* also supports the division-of-labor argument. It recommends developing high-level technologies, shifting to higher-value-added products, and keeping the production center of these products at home to prevent the hollowing-out effect of FDI. The *Economic White Paper* uses the dynamic random access memory (DRAM) chip industry as an example of shifting to a higher-level industry to avoid hollowing-out effects.[72] However, the case of the same DRAM chip industry can be used to support a counterargument.

The comparison of U.S. and Japanese industrial strategies in electronics during the 1960s and 1970s will make the issue clearer. Most American observers at the time thought that the Japanese might be able to occupy the lower end of the business but would not be able to develop the products that would threaten the major components of the product range. Contrary to this expectation, Japanese industrialists targeted the weak spots of competitors first and then used the accumulated production experience in this lower end of the market to attack the higher end later.[73] A good example is the series of product developments of transistor radios, televisions, VTRs, semiconductors, robots, and even high-definition color televisions in recent years.

The major reason for this Japanese success lay in a long-term market and production strategy that emphasized utilizing intraindustry linkages. This strategy was founded on a firm commitment to capital investment at home to upgrade technology and production experience and capabilities. The upgraded technology and the accumulated production experience and capabilities in the lower-level electronics goods led to the

[71] Reich (fn. 42), 21.

[72] Economic Planning Agency (fn. 28), 124–26.

[73] I. C. Magaziner and T. M. Hout, *Japanese Industrial Policy* (Berkeley: University of California Press, 1980), 16.

successful development of higher-level and higher-value-added electronics goods in the later period. For instance, the accumulated production experience and upgraded technology in color television set production proved to be a basis for the success in VTR production, and the success in consumer electronics was a driving force behind the success in the DRAM memory and other semiconductor industries.

The U.S. approach was different. At the time the Japanese color television and semiconductor industries were emphasizing domestic capital investment as a way of enhancing competitiveness and therefore automated their production facilities, American industrialists in the electronics industry depended instead on the offshore production strategy.[74] This strategy locked U.S. companies into the existing state-of-the-art technology and eliminated further need to retool and automate.[75] A study shows that

> Japanese television equipment manufacturers responded to the recession [of 1974–1975] by retooling and automating production possibilities. The U.S. industry's reaction was to cut costs by moving offshore to low-wage economies and locking into then-current designs and production techniques, and to release R and D personnel as a further cost-reduction measure.[76]

In this sense, FDI for U.S. industrialists has become a disincentive for domestic automation.[77] But automation of production facilities in Japan led to the improvement of productivity, product quality, and competitiveness. This was the main reason for their victory over American products in consumer electronics and DRAM chips.[78] For example, the average number of hours required to assemble a twenty-inch color television set in 1977 was about 1.5 in Japan, 3.5 in the United States, and 5.8 in the

[74] Jack Baranson, *The Japanese Challenge to U.S. Industry* (Lexington, MA: Lexington Books, 1981), 100–102, 129–30; Daniel I. Okimoto, Takuo Sugano, and Franklin B. Weinstein, eds., *Competitive Edge: The Semiconductor Industry in the U.S. and Japan* (Stanford, CA: Stanford University Press, 1984), 62–64; James E. Millstein, "Decline in an Expanding Industry: Japanese Competition in Color Television," in John Zysman and Laura Tyson, eds., *American Industry in International Competition* (Ithaca, NY: Cornell University Press, 1984), 112, 114–17, 120–21, 136–38.

[75] Baranson (fn. 74), 101. See also Millstein (fn. 74), 106–7.

[76] World Developing Industry and Technology, Inc., "Sources of Competitiveness in the Japanese Color Television and Tape Recorder Industry," a study for the U.S. Department of Labor (October 1978), 21, as quoted by Millstein (fn. 74), 127.

[77] Baranson (fn. 74), 101. See also Millstein (fn. 74), 115, and Okimoto, Sugano, and Weinstein (fn. 74), 64.

[78] Both manufacturers and purchasers of semiconductors agreed that during the mid- to late 1970s, quality levels delivered by Japanese firms were superior to those delivered by U.S. firms. See U.S. Congress, Office of Technology Assessment, *International Competition in Electronics* (Washington, DC: Government Printing Office, 1983), 230. For a quality test of Japanese and American chips by Richard W. Anderson of Hewlett-Packard, see "Japan Makes Them Better," *Economist*, April 26, 1980, p. 55.

United Kingdom.[79] Also, according to a survey of nearly 200,000 owr of nineteen-inch color television sets sold during the period 1975–19 nine of the fifteen brands—all the Japanese makers—were given r ability ratings of "better than average," whereas of the American brar one was rated "average" and five others "worse than average."[80]

If the Japanese industry had depended as heavily on FDI as had U.S. consumer electronics and chip industries, it would not ha achieved its present level of competitiveness and advanced technologi In short, the transfer of production of lower-level products abroad lea to the loss of grounds for the development of higher-level products ar breaks the intraindustry linkage. The division-of-labor argument co tradicts the core of this long-term strategy. The assertion that hollowing out can be prevented by controlling certain industries or parts of th production process at home argues in favor of following the past experience of the U.S. consumer electronics and chip industries. Some may say that Japanese FDI is different from American FDI because it has evolved under conditions of careful consideration of domestic production and supplemented domestic investments. The recommendation of the *Economic White Paper* that the production center of high-valued products be maintained in Japan is also based on this assumption. But it may have been possible to maintain this supplementary relationship during the first phase of trade-oriented economy because FDI was not yet important to the economy as a whole at that time. The relationship will not necessarily be as easily maintained during the new period of massive and competitive FDI. Once oligopolistic competition in the FDI arena is unleashed so that firms may realize higher profits, it will be extremely difficult to control the flow of FDI by determining which industries should venture abroad and which should remain at home.[81] The Japanese cannot successfully pursue FDI and another, conflicting goal at the same time: the division of labor regarding capital investment among countries operates according to its own logic.

If we briefly expand our scope and incorporate the effect of foreign portfolio investments into our discussion, the picture becomes clearer.[8]

[79] Magaziner and Hout (fn. 73), 26–27.

[80] "19 Inch Color TVs," *Consumer Reports* (January 1981), 34.

[81] Knickerbocker emphasized the dynamic aspect of oligopolistic competition leading t foreign direct investments instead of focusing on the static aspect, as did most oligopol explanations. See F. T. Knickerbocker, *Oligopolistic Reaction and Multinational Enterpris* (Boston: Division of Research, Graduate School of Business Administration, Harvard Uni versity, 1973), esp. chaps. 2 and 3.

[82] Compared with the British and American experiences, the Japanese case shows an im portant difference in foreign investment behavior. In the British case, foreign portfolio ir vestments by financiers were predominant. In the American case, FDI by industrialists wa predominant. In the Japanese case, however, foreign portfolio investments were still predom inant in the late 1980s, and an important portion was being made by industrialists. For ir

For instance, the existence of easy income sources abroad diverted industrialists' resources and energy from their main line of production. In 1985 more than half of Japan's purchases of foreign bonds came from companies investing their profits.[83] According to a survey of major firms by Mitsubishi Bank, the companies' operational profit based on production activity increased only 2.6 times during the past decade, whereas their pretax profits, which also included financial account earnings, rose by a factor of 5.6.[84] The survey estimated that funds diverted for *zai-tech* (sophisticated money management) purposes by 1985 totaled 10–12 trillion yen, of which 4 trillion yen was added during 1985 alone.[85] Table 5 indicates (1) the profits from nonproductive activities for the five companies that marked the highest profits from *zai-tech* in 1986 and (2) the relationship of those profits, expressed as a percentage, to the operational profits (from mainline production).

For many firms, financial revenues became more important than revenues from mainline production. These Japanese firms regarded *zai-tech* as a key profit base and set up specialized financial subsidiaries. However, this kind of financial speculation usually depresses entrepreneurship in the top management of many companies and brings a loss to the national economy by diverting a considerable amount of money from investment in physical assets and R & D, the most important factors of productivity growth. Even those industrial managers who were involved in *zai-tech* admitted the negative effect of this speculative fund management. For instance, the president of Olympus Optical Company origi-

TABLE 5
PROFITS FROM NONPRODUCTIVE ACTIVITIES

	Amount of Profit (billions of yen)	*Nonproductive Profits as Percentage of Mainline Production Profits*
1. Toyota	159	32.6
2. Nissan	127	107.1
3. Matsushita	115	64.6
4. Mitsubishi	46	58.3
5. Sony	39	111.7

SOURCE: Wakō Kezai Kenkyūsho, *Shoken Tōshi* [Stock investment] (July 1987), 36.

stance, in 1982, 34% of the total investment in foreign securities was accomplished by private nonfinancial corporations. See Economic Planning Agency, *Keizai Hakusho 1986: Kokusaiteki Chōwa o Mezasu Nippon Keizai* [Economic white paper 1989: Japanese economy aiming at international harmony] (Tokyo: Ōkura-shō Insatsukyoku, 1986), 565. This is a new phenomenon in the history of foreign investments, one that has not yet attracted much attention.

[83] See "Japan's Foreign Investment: Home Is Where the Hurt Is," *Economist*, November 9, 1985, p. 94.

[84] "'Zai-tech' Boom," *Japan Economic Journal*, March 29, 1986, p. 6.

[85] Ibid.

nally thought that finance-oriented management tended to discoura efforts to produce high-quality products. However, financial profits his firm accounted for about 37 percent of recurring pretax profits in tl period from November 1985 to October 1986.[86] Sixty-four percent of tl 776 companies surveyed by the Tokyo Chamber of Commerce and I dustry felt that *zai-tech* was an undesirable option as far as ordinary co porate activities were concerned.[87]

The case of Tateho Chemical Industries Company was typical of ho some Japanese industrial firms, instead of trying to maintain their con petitiveness in mainline production, became heavily involved in financi speculation, which eventually proved disastrous.[88] Though the intensi of the *zai-tech* boom has diminished somewhat since the stock mark crash of October 1987 and the recent expansion of domestic demand, tl tendency of industrial companies to depend on financial speculation wi not quickly disappear.

Rapidly expanding foreign direct (and portfolio) investments wi bring handsome profits in the immediate future. These net foreign ear ings will increase and continue to make a great contribution to Japan balance of payments in the future, obviating the need for trade su pluses.[89] The economy, however, will be increasingly dominated by f nancial services. A recent report predicted that the boom in the servic sector would create 10 million new jobs over the next fifteen years an that the manufacturing sector would lose 2.6 million jobs.[90] If this tren continues, the Japanese economy will gradually become a rentier eco omy, as did such predecessors as Great Britain before 1914 and tl United States since the 1970s. More and more talented university grad uates in Japan are applying for jobs in the financial rather than the e gineering sector, just as such graduates have flocked to Wall Street i the United States.

Beyond FDI

Though not yet a hegemonic power, Japan has entered upon the statu of a quasi hegemon in economic dimension. As such, the path of its eco omy has begun to mark a trajectory very similar to that of two earlie

[86] "Companies Increasingly Turn to 'Zaitech' to Combat Profit Fall," *Japan Economic Jou nal*, June 6, 1987.

[87] *Japan Times*, February 5, 1988.

[88] *Wall Street Journal*, September 28, 1987, p. 25, and *Japan Economic Journal*, Septemb 19, 1987, p. 6.

[89] Profits from investments abroad are projected to be $400 billion by 1990. See "Settir Sun? Japan Inc.'s Rise Holds the Seeds of Its Fall," *Washington Post*, June 22, 1986, p. C1.

[90] *Business Week* (fn. 40), p. 58.

hegemons, Great Britain in the late nineteenth century and the United States after World War II. What lessons can Japan learn from their experiences, and what should be the direction of change? In answering this question, I want to remind the reader of the comment of an observer at the beginning of this century. Arguing that "mal-distribution of consuming power which prevents the absorption of commodities and capital within the country" was the cause of foreign investments, J. A. Hobson asserted that "the struggle for markets, the greater eagerness of producers to sell than of consumers to buy, is the crowning proof of a false economy of distribution. Imperialism is the fruit of this false economy; 'social reform' is its remedy."[91]

Over eighty years later, Robert Gilpin borrowed Marxist language in the following statement: "Unwilling to make the needed domestic reforms, Japanese capitalism therefore required a 'colony' to rid itself of these financial surpluses. The Japanese found this 'vent for surplus' in an America experimenting with Reaganomics."[92]

Putting aside the question of whether these two scholars' theoretical perspectives have more relevance than those of economists in explaining the cause and effect of Japanese FDI, it is clear that there exists a parallel between the circumstances described by these comments and the current Japanese situation. This parallel is readily apparent in the ironic imbalance between the standard of living of the Japanese and the country's huge accumulation of capital, as I have shown. This is why stimulating domestic demand and investing in the domestic infrastructure through institutional reforms is a better strategy than that of FDI. Stimulation of domestic demands will improve the domestic standard of living (which is, after all, the ultimate goal of economic activities), and investments in the infrastructure will lead to further productivity growth in the future. By contrast, continuing FDI, as neoclassical economic analyses show, will exacerbate the unequal income distribution between labor and capital.

What will happen if there is no change? An article in the *Washington Post*, headlined "Setting Sun? Japan Inc.'s Rise Holds the Seeds of Its Fall," concludes as follows:

> But if its economy is increasingly dominated by financial services, that could subvert the very basis of Japan's industrial success. . . . Unless there are fundamental changes in the nation's economic structure, Japan will be hard pressed to avoid the mistakes that have led to the decline of other great nations.[93]

[91] Hobson, *Imperialism* (Ann Arbor: Ann Arbor Paperbacks, University of Michigan Press, 1983), 85, 88.

[92] Gilpin, *The Political Economy of International Relations* (Princeton: Princeton University Press, 1987), 330.

[93] "Setting Sun?" (fn. 89).

PRIDE OF PLACE
The Origins of German Hegemony

By BRUCE BUENO DE MESQUITA*

THEORIES of power-transition, or hegemonic, war assume thai tem-transforming wars must be large, especially costly conflict: tween adversaries who hold different views of how the international tem should be run. In this essay, using game theory and a deta historical case analysis, I shall offer an existence proof that system-tr: forming wars can be small and that the rivals need not disagree over functioning of the international system: a small event—the Se Weeks' War of 1866—fundamentally changed the international or by providing the foundation for German hegemony on the Europ continent.

To demonstrate the argument, I begin with the conditions stipula by hegemonic war/power-transition theorists as necessary for a syste transforming war. After presenting a game-theoretic model of inter tional affairs, including war, I identify the conditions that are necess and sufficient for war in this game. Having developed the complem tarity between the game theory model and theories of hegemony, I sh that the push toward German hegemony following the Seven We War conforms to the expectations of that model. Yet, the Seven We War has been overlooked by theorists of hegemonic war who conf the consequences of system-transforming wars with the causes and pr imate conditions of these wars.

Hegemonic War and the Status Quo

Theories of hegemonic war can be traced back to Thucydides, M kinder, and Carr; in their modern manifestations and current deb they are clearly rooted in the research of Organski, Kugler, Gilpin, . Kennedy, among others. The essence of the key variants of the theor that differential rates of growth lead some nations to rise in power . others to fall. When the dominant state in an international systen

* I wish to thank David Brown, James Fearon, John Ferejohn, Lewis Gann, Leo Hochberg, T. Clifton Morgan, James Morrow, and Thomas Sargent for their helpful gestions. I also wish to thank Barry Weingast both for his invaluable comments and fo encouragement he offered in this undertaking.

being surpassed in power by a rival that is *dissatisfied* with the international order—the prevailing international status quo—the risk of hegemonic war is great. Such power-transition wars culminate in a restructuring of the international system around the leadership of the newly emergent hegemon.[1]

Questions of who fights in hegemonic wars and the reasons such wars are fought are at the heart of theories and historical evaluations of relations among competing great powers. Organski and Kugler, Gilpin, and Kennedy offer remarkably similar answers to these questions. Organski and Kugler assert that power-transition or hegemonic wars arise because great powers "desire to redraft the rules by which relations among nations work."[2] Gilpin says of such disputes that

> the war determines who will govern the international system and whose interests will be primarily served by the new international order. The war leads to a redistribution of territory among the states in the system, a new set of rules of the system, a revised international division of labor, etc.[3]

Organski and Kugler maintain that the aggressor in a power-transition war "will come from a small group of dissatisfied strong countries" and that "it is a general dissatisfaction with its position in the system, and a desire to redraft the rules by which relations among nations work, that move a country to begin a major war."[4] In a similar vein, Gilpin argues that "the fundamental issue at stake is the nature and governance of the system":

> The disequilibrium in the international system is due to increasing disjuncture between the existing governance of the system and the redistribution of power in the system. . . . From the perspective of dominant powers, the costs of maintaining the international status quo have increased. . . . From the perspective of rising powers, the perceived costs of changing the international system have declined relative to the potential benefits of doing so. . . . Thus, in accordance with the law of demand, the rising state, as its power increases, will seek to change the status quo as the perceived

[1] Thucydides, *History of the Peloponnesian War*, trans. Rex Warner (Middlesex: Penguin Books, 1954); Halford Mackinder, *Democratic Ideals and Reality* (New York: W. W. Norton, 1962); Edward Hallett Carr, *The Twenty Years' Crisis, 1919–1939* (London: Macmillan, 1951); A.F.K. Organski, *World Politics* (New York: Alfred Knopf, 1958); A.F.K. Organski and Jacek Kugler, *The War Ledger* (Chicago: University of Chicago Press, 1980); Robert Gilpin, *War and Change in World Politics* (Cambridge: Cambridge University Press, 1981); Paul Kennedy, *The Rise and Fall of the Great Powers* (New York: Random House, 1987). See also Charles Doran and Wes Parsons, "War and the Cycle of Relative Power," *American Political Science Review* 74 (December 1980), 947–65; George Modelski, *Long Cycles in World Politics* (Seattle. University of Washington, 1987); William R. Thompson, *On Global War* (Columbia: University of South Carolina Press, 1988).

[2] Organski and Kugler (fn. 1), 23.

[3] Gilpin (fn. 1), 198.

[4] Organski and Kugler (fn. 1), 19, 23.

> *potential benefits begin to exceed the perceived costs of undertakin change in the system.*[5]

Apparently, conflicting views of the international status quo are hyp esized to be a key casus belli in system-transforming wars.

Students of hegemonic or power-transition wars also agree broad the specification of criteria for identifying the events that could tr form the international system. As Gilpin notes, "the most important sequence of a hegemonic war is that it changes the system in accord with the new international distribution of power"; therefore, "such a . . . becomes total and in time is characterized by participation by all major states and most of the minor states in the system. . . . For reason, hegemonic wars are unlimited conflicts; they are at once polit economic, and ideological. . . ."[6] Organski and Kugler observe that wars must include major powers as participants in each opposing co tion, and that they must make "an all-out effort to win" so that number of battle deaths reache[s] higher levels than in any prev war."[7]

The transformation of the international order through a large, blo war is believed to be a consequence of hegemonic decline.[8] Yet sys transformation can occur without such a conflict, or even when su transformation was not the objective of relevant disputants. Moreov system transformation may fail to materialize even though it was objective. A somewhat misleading view of such system-transforn wars may arise because of a tendency to study only those cases in w a large war was motivated, at least partially, by a desire to "redraft rules" of the international system. By selecting cases only from these ter wars, researchers risk misstating the conditions that are either essary or sufficient for such events. In order to delineate the neces and sufficient conditions under which system-transforming wars can cur more fully, I propose a game-theoretic model as an addendum theories of hegemony.

The International Interaction Game and the Status Quo

International politics are, of course, so diverse that no two interact are likely to be exactly the same. Yet several essential features are c mon to the development of any relationship between states. These

[5] Gilpin (fn. 1), 198, 186–87.
[6] Ibid., 198–99.
[7] Organski and Kugler (fn. 1), 46.
[8] For similar views of the causes of systemic wars, see, for instance, Manus Midlarsk *Onset of World War* (Boston: Unwin Hyman, 1988); Thompson (fn. 1).

tures relate to sequences of decisions that encourage either friendly or hostile relations. To describe them, I begin with the assumption that each nation can be treated as a unitary actor. In the application of this assumption, I recognize that the metaphoric unitary actor is responsive to domestic political considerations as well as to calculations of support from its allies and of support for its enemies from *their* allies.

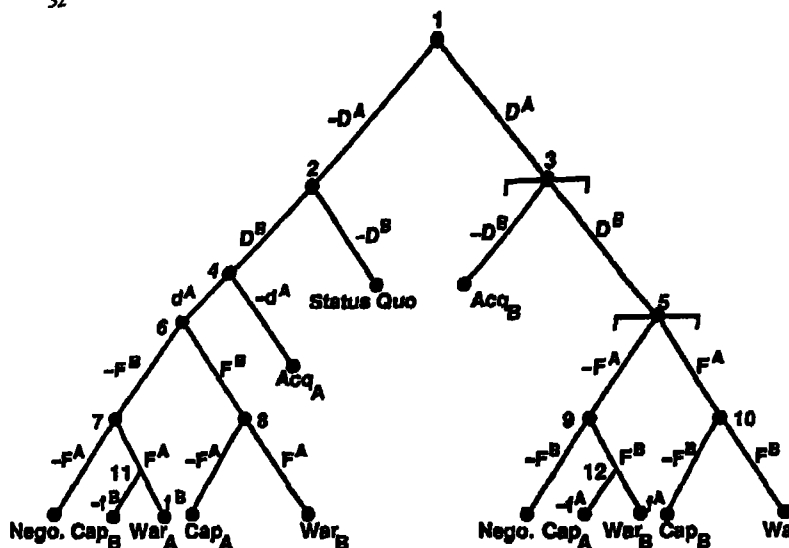
International interactions are assumed to arise in a context that provides some state with an opportunity to take the initiative in shaping relations—an initiative that may govern the future development of events. States shape their relationships through the selection of strategies. Different combinations of strategies result in different political outcomes. If, for instance, two states exchange demands and each state uses force to try to accomplish its goal, a war ensues. If the interaction between two states involves no demands, each has elected to live with the status quo. The selection of strategies is a function of the value that a state attaches to alternative outcomes and the beliefs it holds with regard to how its adversary will respond to its strategic decisions.

Let us assume that decision makers respond to circumstances by making the choice that they believe maximizes their expected utility from that stage of the game onward. They cannot precommit themselves to one or another future course of action—in the largely anarchic system of international relations no binding authority exists to enforce agreements against the participants' own will—but they can act in anticipation of their opponents' choices. That is, we assume subgame perfection in which forward-looking decision makers contemplate the consequences of their current strategic choices for an entire sequence of interactions.[9]

Let us further assume that state A has the opportunity to initiate an interaction; it can choose to make a demand (D^A) or not make a demand ($\sim D^A$) of another state. The demand may be about anything. Once A has made its move, as shown in the extensive form game depicted in Figure 1, state B has the opportunity to select a course of action. Using superscripts to name the actor selecting a strategy, we can say that if the initial sequence is $\sim D^A, \sim D^B$, the outcome is the maintenance of the status quo. If the sequence is $D^A, \sim D^B$, then B is said to acquiesce to A's demand. If A forgoes the opportunity to initiate a demand and allows B to initiate ($\sim D^A, D^B$), then A, by not making a counterdemand ($\sim d^A$), can acquiesce to B's demand ($\sim D^A, D^B, \sim d^A \rightarrow$ Acquiescence to B's demand), or A can make a counterdemand ($\sim D^A, D^B, d^A \rightarrow$ Crisis). Similarly, if the initial sequence of strategic choices were D^A, D^B, there would also be a crisis by

[9] Reinhardt Selten, "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory* 4, no. 1 (1975), 25–55.

FIGURE 1
INTERNATIONAL INTERACTION GAME

A's second move. Failure by both parties to abide by the status quo a failure by either party to acquiesce to the other's demand results in crisis.[10]

In a crisis, a state may choose to escalate the dispute further by usi force, or it can attempt to defuse the situation by offering to negotia abiding by the status quo ante is no longer an option.[11] If there is an off to negotiate, the adversary can reciprocate by not using force, or it c exploit the offer to bargain by attacking. Any time a state escalates dispute by using force, it can expect one of two responses. The defend can capitulate to the attacker's demands, thereby cutting its battlefi losses, or it can retaliate, thereby escalating the dispute further so that becomes a war.

The bracketed portion of Figure 1 (node 5) highlights the cri subgame. In the crisis subgame, A faces the choice of offering to nego ate by not initiating the use of force ($\sim F^A$) or of escalating the dispute

[10] David Lalman, "Conflict Resolution and Peace," *American Journal of Political Science* (August 1988), 590–615; Robert Powell, "Crisis Bargaining, Escalation, and MAD," *Ameri Political Science Review* 81 (September 1987), 717–35; James Morrow, "A Continuous-O come Expected Utility Theory of War," *Journal of Conflict Resolution* 29 (September 19 473–502; idem, "Capabilities, Uncertainty and Resolve: A Limited Information Model Crisis Bargaining," *American Journal of Political Science* 33 (November 1989), 941–72.

[11] Lalman (fn. 10).

using force to back up its demand (F^A). If A uses force, B must choose between capitulating to A's first strike (F^A,F^B). In the latter case, the strategy sequence $D^A,D^B,F^A,^A,\sim F^B$) or striking back by using force itself (F^A,F^B). In the latter case, the strategy sequence D^A,D^B,F^A,F^B results in a war initiated by state A.

If A offers to negotiate ($\sim F^A$), then B's choices can lead either to negotiations ($\sim F^A,\sim F^B$) or to escalation ($\sim F^A,F^B$). If B selects the escalatory path, then A must make a final strategic determination: to capitulate ($\sim f^A$) to B's enforced demand ($\sim F^A,F^B,\sim f^A$) or to retaliate ($\sim F^A,F^B,f^A$), resulting in a war initiated by B.

As is evident from Figure 1, international interactions can culminate in eight different generic outcomes:

> status quo; acquiescence to A's demand; acquiescence to B's demand; negotiation; capitulation to A; capitulation to B; war initiated by A; and war initiated by B.

Each of the possible outcomes can be characterized in terms of the expected value of the outcome to each player. The Appendix displays the assumptions that constrain the possible sequence of preferences over the game's otucomes.[12] With the assumptions taken into account, there are 52 possible orderings of the expected payoffs of the 8 outcomes of the game. Therefore, 2,704 unique combinations of orderings of expected outcomes are possible.

Most of the admissible orderings imply negotiations or the preservation of the status quo as the subgame perfect equilibrium outcome, but some orderings yield the expectation of war. The concern here is with the possibility of a system-transforming war when the status quo is highly valued, a condition explicitly ruled out by theories of hegemonic war or of power transition, and therefore a contradiction of those theories.

To establish that war is possible even though the status quo is highly valued, we must identify the conditions that are necessary and sufficient for war to be the equilibrium outcome of the international interaction game; then we can explain how uncertainty resulting from changes in the power of the rivals may affect the likelihood of war. This section concludes with a discussion of how the costs of war are linked to the value of the goals being pursued.[13]

The game implies the following theorem:

[12] The algebraic representation of the game's payoffs is found in Bruce Bueno de Mesquita and David Lalman, *War and Reason* (New Haven: Yale University Press, forthcoming).

[13] See ibid. for a more technical treatment of the relationship between the costs of war and the equilibrium conditions for war.

War Theorem: War initiated by state A is the equilibrium outcome of international interaction game if state A prefers to initiate a war ra than acquiesce to its adversary's demands; state B prefers to try to f the adversary to capitulate rather than to offer a negotiated settlemen believes A would capitulate rather than retaliate if *B* strikes A; and prepared to strike back at its adversary rather than give in to its demanc

The theorem indicates that states are inclined to strike first if believe that they risk exploitation by offering to negotiate. The attached to the status quo does not enter into the theorem at all. Co quently, whether a war is system transforming or not, the argumen power-transition/hegemonic war theorists to the effect that dissatis tion with the international status quo is a necessary condition is con dicted by the international interaction game.

Although the above theorem is not consistent with hegemonic theories regarding the status quo, it is consistent with important asp of those theories. In particular, let us recall that such theories conte that the risk of system-transforming war is heightened when a rival creases in power sufficiently to challenge the hegemon. As the hegemc preponderance is eroded, uncertainty increases regarding each acto probability of defeating the other in a confrontation. In the face of su uncertainty, the decision makers' expectations about each other's fut actions are critical.

A central source of uncertainty concerns whether the adversary is *lieved* to prefer to retaliate if coerced. Suppose, for instance, that state possesses private information that encourages its leadership to beli that it could win a war even if B struck first. Suppose B were uncert of A's expectations; specifically, assume that B was sufficiently confid that it could still deter A. Then B would respond to an offer of negot tions by trying to coerce A into capitulating to B's demands. If A anti pates this response by B, then A will seize the initiative and attack B fi A's growth in power is, in this example, the source of the uncertainty suggested by hegemonic war/power-transition theorists, and is the cr cial ingredient pushing the rivals toward war even though they may satisfied with the status quo.[15]

[14] For proofs of the theorem, see ibid., and Bruce Bueno de Mesquita and David Lalm "The Road to War is Strewn with Peaceful Intentions," in Peter Ordeshook, ed., *Model Strategic Choice in Politics* (Ann Arbor: University of Michigan Press, 1989), 253–66.

[15] For ease of presentation, assume that state *B* is risk-neutral or risk-averse. Let B uncertain about its own and state A's subjective probability of success. Let B believe w probability β that A's probability of success is low enough that A will not retaliate if attack Then, using the notation explained in the Appendix, whether A actually prefers to retal or not, *B will try to exploit* A if $\beta > [\alpha^B + \phi^B] / [\alpha + (1 - P^B)(G^B - L^B)]$; and A, anticipat this exploitation, will prefer to start a war. A formal proof of this argument can be found Bueno de Mesquita and Lalman (fn. 12).

The war theorem reveals special characteristics of wars in which the rivals are satisfied with the status quo. In particular, a partial equilibrium condition of the war theorem is that the probability of success must be greater than a positive function of the ratio of the expected costs to the expected stakes, with the stakes defined as the difference between the utility attached to obtaining one's objectives and the utility attached to giving in to the rival's demands. If the status quo is highly valued *by both sides*, then it is likely that the magnitude of either side's demands does not deviate much from the value of the status quo, so that the stakes are small. Under those circumstances, the condition just described can only be satisfied if the expected costs are small. Thus, when the status quo is highly valued, the war will be fought only if the *expected* costs are small. And, when the battlefield costs rise above expectations without a concomitant increase in the stakes, it seems likely that the war will end.

The partial equilibrium condition, then, reveals that a war fought over small changes in the status quo must necessarily be expected to be a small war. Of course, as argued earlier, a small event may have profound consequences for the international order. For a demonstration, I turn to an assessment of the Seven Weeks' War.

The Seven Weeks' War: Pride of Place and the *Status* Quo

Prior to the war, an Austro-Prussian crisis had arisen because each side made demands regarding the disposition of the Elbe Duchies, and neither was willing to accept the other's wishes. Recall that, in 1864, Prussia and Austria fought as allies against Denmark during the Second Schleswig-Holstein War. As a consequence of that conflict, these two preeminent German states secured control over the disputed duchies of Schleswig and Holstein. Almost as soon as the war ended, internecine struggles ensued between Prussia and Austria over the governance of the captured territories. At stake for Prussia was access to a port for its navy, enhanced prestige among the lesser German states, and the demonstration that Prussia was no longer appreciably weaker than Austria. At the same time, Prussia's ambitions *within* Germany were constrained by its domestic sentiments against a fratricidal conflict and by Austria's favored position as the predominant German power throughout the history of the Germanic Confederation.

The nascent Austro-Prussian conflict that surfaced after their joint success against the much weaker Danes was brought to an apparent end by late 1865 with the signing of the Treaty of Gastein. The treaty accom-

modated several of Prussia's aspirations while protecting and preservi Austria's prestige and preeminence within the German Confederati Yet, in a few short months it became apparent that the treaty was I working. The immediate issues revolved around Bismarck's conce that Austria's right to administer Holstein was tantamount to *sovereig* over the territory. Such an outcome—or even the right of the Germ Diet to decide Holstein's fate—frustrated Bismarck's immediate a long-term ambitions within the Confederation. To rectify the situati from its perspective, Prussia sought authority over Holstein, in part violation of the Treaty of Gastein. Austria sought to extend its own cc trol over Holstein and to restrict Prussian aspirations further by turni the resolution of the Duchy's sovereignty over to the German Diet body more under Austria's domination than under Prussia's. This, t was a violation of the Treaty of Gastein.[16]

The central demand of Prussia amounted to a quest for pride of pla within the leadership of Germany. As *The Economist* summarized t dispute on the eve of the war, "the demand, therefore, is simply that t minor States [of the German Confederation] shall follow Prussian le instead of Austrian lead." Likewise, Austria's posture amounted to desire to cling to the privileged position granted to it by the Congress Vienna—to hold on to its own pride of place as the leader of Germar *The Economist* continues:

> Tradition counts for a good deal among nations, and the German tradition is that the leadership of Germany belongs *de jure* to the "Kaiser," that the King of Prussia is an overpowerful feudatory.... This demand [of Prussia] is, however, an assertion that the feudatory intends to rule, and ... will be resisted to the death by Austria, not only because it is inconvenient, not only because, like the claim to the Elbe Duchies, it greatly increases Prussian power, but because it terminates her own career as a German state, a career of which ruler and people are very proud, and also very jealous.[17]

Prussia's ambitions before the War of 1866 seem to have been rel tively modest in the greater European scheme of things. Bismarck's e ergies before the war had largely been focused on securing his and t

[16] The Treaty of Gastein was supposed to settle the remaining questions of sovereig over Schleswig and Holstein that resulted from the Austro-Prussian defeat of Denmark 1864. Article 1 of the treaty stipulates that

> the common right obtained by the high contracting parties by Article 3 of the Treaty of Vienna of the 30th of October, 1864, is transferred, as respects the Duchy of Holstein, to his Majesty the Emperor of Austria, and as respects the Duchy of Schleswig, to his Majesty the King of Prussia, *without prejudice to the continuation of these rights of both Powers to the whole of both Duchies* [emphasis added].

See Alexander Malet, *The Overthrow of the Germanic Confederation by Prussia in 1866* (Lc don: Longmans, Green, 1870), 106–10.

[17] *The Economist*, June 16, 1866, p. 699.

king's political authority in the face of a reform-minded parliament and an obstreperous diet, rather than on grander efforts toward a unified Germany.[18] As Simon notes, many believed that "Bismarck had launched the war as a means of escape from his domestic difficulties," and that, while the war culminated in *kleindeutsch* unification, "*Bismarck did not intend it as such*."[19] Indeed, before the war Bismarck's ambitions regarding greater Germany appear to have been so limited that he seemed to be prepared to forgo the defeat of Austria in exchange for securing his limited goals regarding Holstein. Crankshaw comments that "there was a moment when it looked as though Austria might be ready to sell Holstein to Prussia, as she had sold Lauenberg; Bleichröder was active in raising funds for this; Bismarck was ready for it." Further, "had Austria got rid of Holstein for cash and then sold Venetia to Italy, which she could also have done at this time, the face of history would have been changed. For Bismarck would have had no excuse to go to war and no Italian ally if he did so."[20]

The demands of Prussia and Austria created an internal German crisis by casting doubt on the Treaty of Gastein. The crisis arose when Austria and Prussia faced the choice between negotiating a peaceful settlement of the dispute or escalating it by resorting to arms. Central to the argument here is the realization that Prussia and Austria valued the *international* status quo highly. Still, contrary to the assumptions of the theory of hegemonic war, they reached the crisis stage and waged a war of such importance that it transformed the international system.[21]

Recall that a necessary condition for war is that the challenger values initiating war more highly than it values offering to negotiate if there is a significant risk that the hegemon (Austria) will exploit the offer. Recall also that the risk of this situation arising would be heightened if the challenger's (Prussia's) power were underestimated. That these were the circumstances as perceived by Prussia is suggested by the fact that

> agitation continued in Holstein, and matters arrived at a crisis which might easily lead to either peace or war. It was necessary to consider whether a lasting and sincere friendship with Austria might be obtained

[18] See, for instance, Theodore S. Hamerow, ed., *The Age of Bismarck: Documents and Interpretations* (New York: Harper & Row, 1973); Hajo Holborn, "Bismarck's Realpolitik," *Journal of the History of Ideas* 21 (January–March 1960), 84–98; and W. M. Simon, *Germany in the Age of Bismarck* (London: George Allen & Unwin, 1968).

[19] Simon (fn. 18), 30 and 29 (emphasis added).

[20] Edward Crankshaw, *Bismarck* (London: Macmillan, 1981), 197, 200.

[21] For the characterization of crises, see Lalman (fn. 10); Powell (fn. 10); Steven Brams and D. Marc Kilgour, "Are Crises Rational? A Game-Theoretic Analysis," presented at the World Congress of the International Political Science Association, Washington, DC, August 28–September 1, 1988; Jeffrey Banks, "Equilibrium Behavior in Crisis Bargaining Games," photo dup. (1989), University of Rochester.

> by further concession, not in this special case alone, but in the whole Holstein question, or whether it was the policy of Vienna here, as elsewhere, to suppress Prussia, and to prevent her free development in Germany by means of the press, the elements of revolution, the power of the Confederation, overtures for an alliance with France; in fact by every possible means.[22]

According to this account, the Prussian leadership was worried about being exploited by Austria, or about being suppressed by every means possible. This perspective is reinforced by Crankshaw's account. He notes that the Austrians, still believing that they possessed superior strength, were not prepared to negotiate with Bismarck in good faith. Thus, he reports that

> in January 1866 Austria was given what looked like another chance—by Bismarck himself. ... [S]urrender to us the primacy of North Germany and we will help you to reconquer Lombardy. But Francis Joseph refused to consider the matter: he still believed he could manage a two-front war if only Russia and France kept out of it.[23]

Apparently Bismarck had reason to believe that an offer of concessions to Austria would only invite exploitation.[24] He also had reason to believe that Austria underestimated Prussia's ability to wage a successful war. As suggested in the game theory example based on uncertainty, Bismarck possessed information about Prussian power that was not available to the Austrians. For example, he and Moltke knew more about the benefits of the needle gun and about Prussia's superior skills in utilizing the railroads. The Austrians continued to rely on traditional approaches to warfare. With regard to the needle gun, for instance, "the overwhelming weight of opinion at regimental levels in the Austrian army continued to regard the rifle as a glorified pike."[25] Thus, given Austria's tendency to exploit adversaries who were perceived to be weak and given Bismarck's (and Moltke's) higher valuation of Prussia's chances of success, the Prussians chose war over further efforts at negotiations.

In examining the prewar view of the *international* status quo held by

[22] Department of Military History of the Prussian Staff, *The Campaign of 1866 in Germany*, trans. Colonel Von Wright and Captain Henry Hozier (London, 1907), 3. A similar sentiment regarding the Austrian penchant for exploitative behavior in negotiations is reflected in A.J.P. Taylor, *The Habsburg Monarchy, 1809–1918* (Chicago: University of Chicago Press, 1976), 123–38.

[23] Crankshaw (fn. 20), 201.

[24] See A.J.P. Taylor (fn. 22), 133, for an illuminating discussion of Francis Joseph's treatment of the Croats in negotiations in December 1866; even *after* the Seven Weeks' War, Francis Joseph continued to use coercion in dealing with rivals he perceived to be much weaker.

[25] Dennis E. Showalter, *Railroads and Rifles: Soldiers, Technology and the Unification of Germany* (Hamden, CT: Archon Books, 1975), 52–73, 105–39; quote at 123.

Prussia and Austria, we must carefully separate their disagreements *within* the confines of the German state system from their views of how the international order ought to be functioning. The theory of hegemonic war and power transition emphasizes the value that the rivals attach to the international order, namely, the existing rules and practices that define the international status quo. These theories do not emphasize the local status quo between particular powers. Did Prussia and Austria differ regarding the rules by which the international system ought to have operated? The evidence suggests that they did not.

Austria's external policy, like Prussia's, was hostile to liberal reformers and democratic sentiments. Both Prussia and Austria were anxious to keep France's ambitions in check, as well as to counterbalance Russia in the East. In fact, Prussia and Austria, in addition to having jointly waged war against Denmark in 1864, remained allied with one another almost continuously from 1815 until the dissolution of the Habsburg empire at the end of World War I. Even the Seven Weeks' War only produced a brief interruption in their alliance of mutual defense. Bismarck strove to maintain good relations with Austria, even to the point of opposing Moltke's and King Wilhelm's desire to march on Vienna immediately after the victory at Königgrätz.[26]

In general, the interests of Austria and Prussia outside of Germany had advanced jointly during most of the nineteenth century prior to the 1866 war, even though the two great German states engaged in frequent internecine squabbles. Austria's centrality to the maintenance of the European international order is well documented. As Kennedy notes, "the general peace which prevailed in Europe for decades after 1815 was due chiefly to the position and functions of the Habsburg Empire."[27] In terms of the external affairs on which the power-transition/hegemonic war theories focus, there is little basis for arguing that Prussia and Austria were dissatisfied with the status quo.

In their bilateral relationship, the considerable strains between Austria and Prussia were *not over the international order*, but over pride of place in the German Confederation: the struggle for leadership within Germany. The Austrians presided over the German Diet in Frankfurt and steadfastly refused to share this privilege with the Prussians. To be sure, it was little more than a privilege. As the British envoy to the Diet ob-

[26] Ibid.; Otto Pflanze, ed., *The Unification of Germany, 1848–1871* (New York: Holt, Rinehart & Winston, 1968); Simon (fn. 18); Alan Palmer, *Bismarck* (London: Weidenfeld & Nicolson, 1976); Hamerow (fn. 18); Friedhelm Grützner, *Die Politik Bismarcks, 1862 bis 1871 in der deutschen Geschichtsschreibung* [The politics of Bismarck: 1862 to 1871 in German historiography] (Frankfurt am Main: Verlag Peter Lang, 1986).

[27] Kennedy (fn. 1), 163.

served, "In all matters upon which the two great Powers were agreed, the Diet's proceedings were marked by the most commendable activity and promptitude; if they differed ... the game of official delays and chicanery of all kinds was played to the utmost."[28] Still, it is likely that war would have been averted had the Austrians agreed to shared leadership. Palmer, for instance, reports that,

> if Mensdorff succeeded in calming the military party around Francis Joseph, there was every possibility of a change of front by Austria, as at Gastein the previous summer. Bismarck was, in consequence, determined to have ready an alternative policy which would provide for rapid Austro-Prussian reconciliation should Francis Joseph abandon the traditional Habsburg pretensions to dominate Germany.... Bismarck ... was prepared for a dramatic change in policy right up to the moment the first troops crossed the frontier."[29]

After the war, when Bismarck had removed all beliefs that Prussia was a second-rate power, he again pursued peaceful relations with Austria. Following Prussia's victory, and the establishment of the *kleindeutsch* policy that expelled Austria from Germany, "Bismarck wished to preserve Austria as she was in 1866—defeated, but still German. . . ."[30]

Austria and Prussia disagreed little regarding the operation of the international order; thus, they failed to satisfy a fundamental condition of the theory of hegemonic or power-transition war. But the expectations of those theories were satisfied in other important respects. The power-transition and hegemonic war perspectives hypothesize that differences in rates of growth produce necessary conditions for war. In particular, if the weaker "challenger" is rising to equal and overtake the declining hegemon, a fundamental condition for a power-transition war is satisfied. As Figures 2a and 2b make clear, there was a power transition from declining Austria to rising Prussia. Figure 2a, using pig iron production as an indicator of economic performance, shows that Prussia's share of the great powers' production rose to equal Austria's production around 1837; it remained roughly equal for another ten years, and then overtook Austria's significantly during the next two decades. By 1866, Prussia's economic capabilities had pulled far ahead of Austria's.

As is often the case, the transition of military power lagged behind the economic transition. During the first half of the nineteenth century, Austria maintained a substantial advantage over Prussia in military might when measured as the countries' respective shares of total military personnel and military expenditures among the great powers. From the

[28] Malet (fn. 16), xxii. [29] Palmer (fn. 26), 111. [30] Taylor (fn. 22), 129.

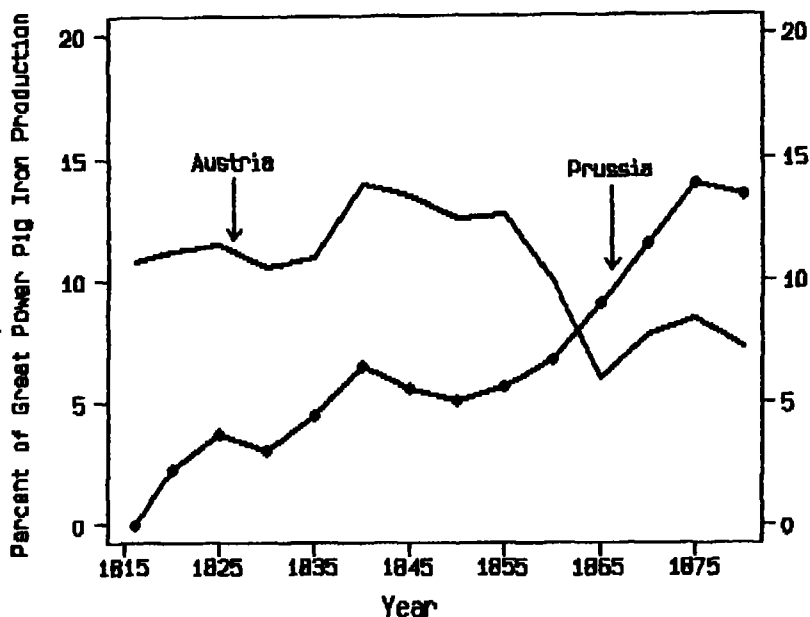



FIGURE 2a
THE AUSTRO-PRUSSIAN ECONOMIC TRANSITION[a]

[a] The data on pig iron were provided by the Correlates of War Project at the University of Michigan.

mid-1850s on, Austria began a precipitous decline in its military capabilities while Bismarck launched Prussia on a campaign of military expansion. According to the indicators used here, the power transition literally occurred in 1866.

The evidence thus far suggests that the critical difference in growth rates postulated by power-transition and hegemonic stability theorists was satisfied in 1866, but that the required disagreement over the *international* status quo was not satisfied. It is possible, however, that there was a widespread *perception* of a sharp disagreement over the international status quo even though, by objective criteria, such differences did not exist. We can never know with certainty what others believed at a given time, but it is possible to develop sensitive indicators that should reflect quite precisely the prewar beliefs about the impact of an Austro-Prussian war on the international status quo.

The cost of money—the money market discount rate—in key financial centers reflects people's *expectations* regarding the future value of that

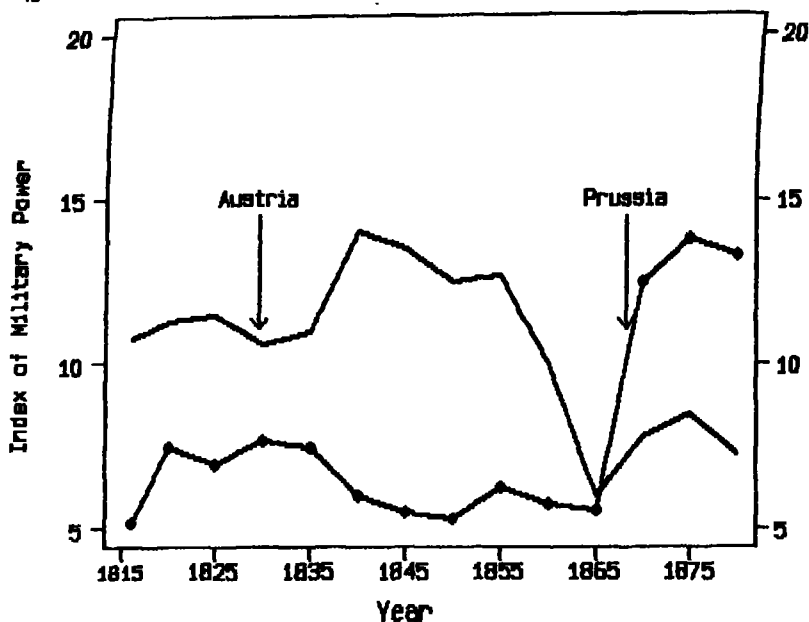



FIGURE 2b
THE AUSTRO-PRUSSIAN MILITARY TRANSITION[a]

[a] The data on military power were provided by the correlates of War Project at the University of Michigan.

money. When, for instance, a government finds it difficult to borrow money, it is forced to raise the rate it pays for the money—the discount rate—to attract lenders. Thus, a rising discount rate for a nation's money reflects a broad base of declining confidence in that nation.

Just as the rise or decline in discount rates reveals information about *expectations*, so do changes in discount rates across countries. If external conditions are expected to affect everyone more or less equally, the money market discount rates for different currencies fluctuate more or less equally; each will respond equivalently to rising or falling fears and uncertainties. But, *if some countries are expected to be differentially affected by events, their rates will rise or fall (depending upon the content of expectations) more than that of other, less affected, countries.*

Figure 3a depicts biweekly observations of the money market discount rate for Berlin and for a European baseline (defined as the average of the discount rates for London and Amsterdam, two key financial centers in the nineteenth century) between January 1863 and January 1865. The

Second Schleswig-Holstein War began in early 1864, with a cease-fire *interrupting the combat for all of May and most of June 1864.* When negotiations failed during the cease-fire, fighting resumed and Denmark was defeated, losing control over the Elbe Duchies. The expectation that Denmark would lose the war, and territory, to the Prussian/Austrian alliance is clearly reflected in Figure 3a. The four small marks on the time line between 1864 and 1865 denote the beginning of hostilities, the interruption of war by the cease-fire (and the beginning of negotiations), the collapse of negotiations (and the resumption of fighting), and the conclusion of the war. The cost of money in London and Amsterdam rose at the outset of the war, responding in part to the general disquiet that war provokes in financial circles. The price of money fell just prior to the cease-fire, and rose again sharply in anticipation of the failed negotiations. At the same time, the price of Prussian money *fell* rapidly once the war got under way; it rose to meet the European base rate only *after* the war ended. The clear separation in money discount rates, and the decline in the price Berlin had to pay relative to other money centers, suggests that the status quo was expected to change in Prussia's favor because of the Second Schleswig-Holstein War.

The mean spread in the money market rates between the London-Amsterdam base and Prussia from January 1, 1863, to the outbreak of hostilities on February 1, 1864, when compared to the period of the fighting (February 1 through July 20, 1864) increased significantly during the war. The t-statistic is 4.42 ($N = 42$), indicating that the events of the war almost certainly influenced the competitive edge gained by Prussian funds.[31]

Figure 3b tracks the same variables as Figure 3a, for the period 1865–1868. The small marks on the time line denote the signing of the Treaty of Gastein (August 1865), the announcement of the Italo-Prussian alliance (April 8, 1866), the mobilization of the Prussian army on May 5 (the Austrians having made their military preparations by early march), the recognition of the inevitability of war by June 16, and the conclusion of the war in late July. The uncertainties provoked by the expectation of war are reflected in the rising discount rate for money. That the great swings in the cost of money were driven by war fears is supported by the accounts in *The Economist* that accompany the weekly reportage of discount rates: "the Bourse has been greatly agitated every day . . . by the dread of war in Germany." Again, "the Bourse has been dreadfully agi-

[31] Comparable indicators for Vienna cannot be utilized as the Austrian discount rate was fixed by the government rather than by auction and was not permitted to fluctuate frequently.

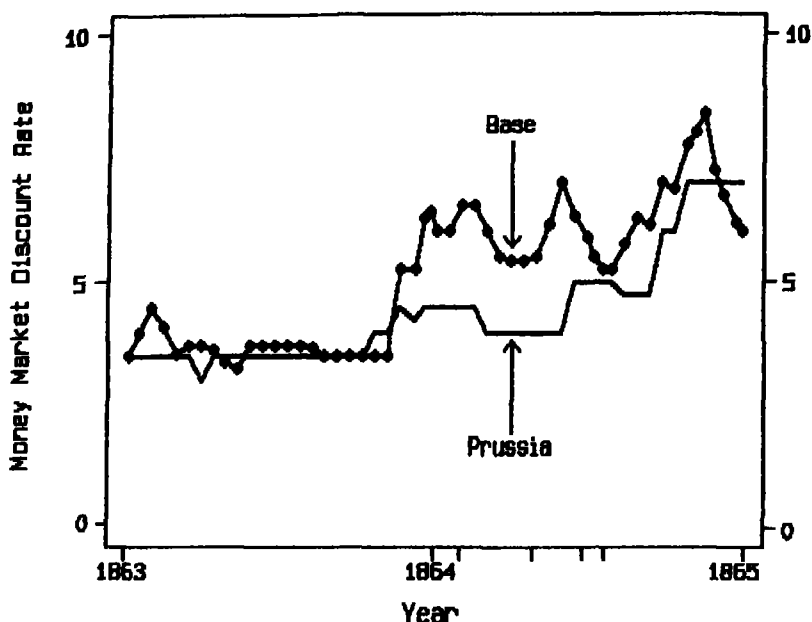



FIGURE 3a
THE SECOND SCHLESWIG-HOLSTEIN WAR, 1864

tated during the past week by the prospect of war, and prices have fallen heavily."[32] Similar reports can be found virtually every week up to the outbreak of hostilities.

Figure 3b demonstrates that, despite the fear of war and its reflection in the rising cost of money, the crisis did not have a substantial differential impact on the expectations concerning Prussia. During the war, however, while the cost of money rose markedly in London, Berlin, and Amsterdam, it rose *more* in Prussia. This reinforces the widely reported observation that Prussia was expected to lose the war.[33] The fact that the price differential between Berlin and the other key financial centers was small supports the belief that the perceived stakes in the war were not very large. It is interesting to note that immediately after the battle of Königgrätz the market responded with a rapid fall in the price of money,

[32] *The Economist*, April 7, 1866, p. 414, and May 5, 1866, p. 535.
[33] Simon echoes the sentiment of many historians when he writes, "it is important to remember that it was by no means a foregone conclusion that Prussia would win; pessimism was widespread in the Prussian camp, and the Austrian government was confident of victory" (fn. 18), 30–31. See also Taylor (fn. 22), 126, regarding expectations from the Austrian perspective, and Showalter (fn. 25), 121, for a general view of Prussian weaknesses.

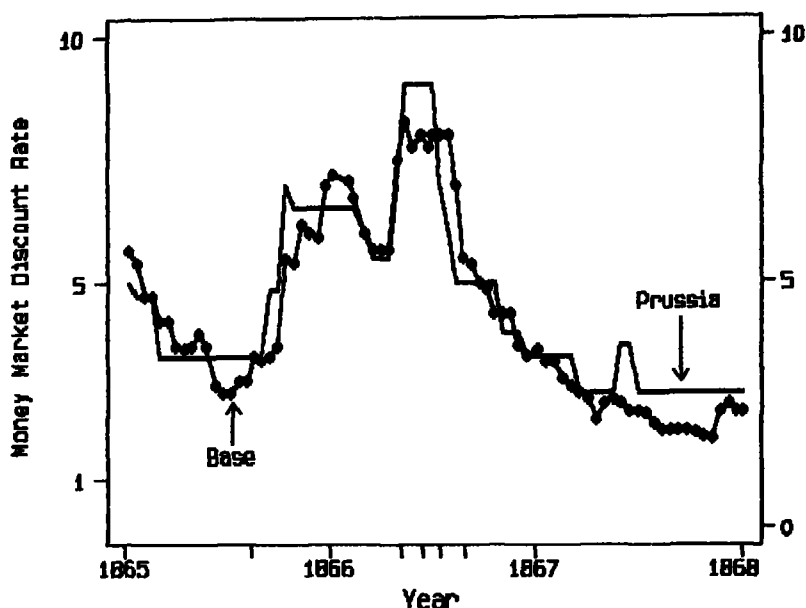



FIGURE 3b
THE SEVEN WEEKS' WAR, 1866

in which Prussia *led* the baseline of London and Amsterdam. Thus, the expectations in the financial markets were updated to take account of the new information revealed on the battlefield—that the market had underestimated Prussia's chance of victory. The prewar fears of postwar inflation or of defaults on money instruments by a defeated Prussia were allayed by Prussia's decisive victory.

Prussia's place in the European international system was not expected, ex ante, to be fundamentally changed by the Seven Weeks' War; this can be seen from a statistical assessment of the difference in the price of money for a significant period prior to the war compared to the period surrounding the war. If the period surrounding the war's crucial events—from the announcement of the Italo-Prussian alliance on April 8 to the end of the war on July 28—had reflected expectations of a fundamental change from the status quo ante for Prussia, the mean *difference* between the Prussian and base discount rate for that period would have been significantly different from the mean difference for the prewar period. If the *international* status quo had not been perceived to be at risk, then there would not have been a significant difference.

Taking the mean difference between Prussia's rate and the average rate for London and Amsterdam from January 1, 1865, to April 1, 1866, as the baseline, I compare this mean to the comparable mean price differential during the events of the crisis and the war. The t-statistic for the difference is -1.054 ($N = 43$), which is *not* significant. The money market data support the argument that Prussia was *not* expected to be differentially affected by the war; the status quo ante—in which Austria was the dominant German state—was expected to prevail.

From this and the preceding discussion we can conclude that, both on objective and on expectational grounds, the Seven Weeks' War was not fought over an alteration of the *international* status quo. The war was not about changing the "rules" of the international order, nor was it fought because one of the antagonists was dissatisfied with that order. Thus, according to the theories of power-transition/hegemonic war, the Seven Weeks' War could not have been a system-transforming war. Yet by the war's end, a major reorganization of Europe came to be expected.

The Seven Weeks' War and System Transformation

During the second half of 1866 and throughout 1867, a sequence of events unleashed forces that were to shape much of the course of European history up to and beyond the outbreak of World War I. *These forces were clearly visible and took meaningful shape well before the outbreak of the Franco-Prussian War, which itself was made possible by the War of 1866.*

Austria, excluded from Germany by the Peace of Nikolsburg, sought new arrangements in a greater Austro-Hungarian Empire to preserve her great-power status. As late as the spring of 1866, just on the eve of a war that was expected to be won by Austria, the Austrian government had been intransigent in its negotiations with Hungary over union. After the war, when Hungary offered the same "deal" as before, Austria no longer had much choice in the matter. As Taylor observes, "The Austria which emerged from the war of 1866 was created by Bismarck as much as the Austria which emerged from the Napoleonic Wars was created by Metternich."[34]

Francis Joseph's willingness to accept the establishment of Austria-Hungary after the Seven Weeks' War was motivated by his desire to secure Austria's position as a great power. For this purpose, he ceded considerable power within the empire to the Magyars and helped to release nationalistic sentiments in Eastern Europe that even today jeopardize stability. According to Michaelis,

[34] Taylor (fn. 22), 126ff.

> While the revolution of 1848 had awakened the national consciousness of eastern European peoples, the Prussian victory over Austria now provided them with the opportunity to fight actively for the realization of their national goals. Königgrätz set in motion the whole eastern realm of nationalities, both inside and outside the monarchy, a movement that was never again to cease.[35]

Taylor makes a similar observation:

> The settlement with Hungary could be challenged only if Czechs and Germans were reconciled, and any real concord among the peoples of Austria, though it would weaken the Magyars, would endanger also the supremacy of the Emperor. . . . Magyar hegemony was the price which Francis Joseph was willing to pay for the preservation of his own power; and since this Magyar hegemony brought the Habsburg Monarchy to destruction, Francis Joseph was the maker of his own ruin.[36]

At least as significant as the forces liberated in Eastern Europe was the war's impact on the balance-of-power system established by the Congress of Vienna. The European order that had generally preserved peace among the great powers of Europe since 1815 was at an end. As Grützner puts it, "a world indeed collapsed, the world of the Congress of Vienna."[37] France and Russia started to press for a new congress to establish order in the emerging Europe and to protect their position and the balance of power.

The collapse of the existing European order had neither been planned by Bismarck nor even been recognized until several weeks after the war ended. The effects of the war remained unclear to Europe's other great powers until Bismarck had achieved the relatively moderate peace terms he desired—a peace that preserved Austrian sovereignty and left open the opportunity for realignment with Austria. While negotiating with the Austrians, Bismarck wrote to Tsar Alexander that "he was not to worry; there would be no drastic change in the map of Europe, since all that Prussia asked was a reform of the Confederation with the expulsion of Austria from Germany, and a final accounting in the matter of the duchies." It was only about a month later, with the war settled and Bismarck's possibilities clearer to him, that France and Russia understood how fundamentally the balance of power had changed.

> Then there was urgent talk of a European congress: if the Continental order established by the Congress of Vienna was to be replaced by a dif-

[35] Herbert Michaelis, "Königgrätz, 1866: Defeat of Liberalism and Universalism," in Pflanze (fn. 26), 106–13.

[36] Taylor (fn. 22), 141. See also Simon (fn. 18) on the formation of Austria-Hungary as a consequence of the Seven Weeks' War.

[37] Grützner (fn. 26), 80. "Tatsächlich stürzte eine Welt ein, die Welt des Wiener Kongresses" (author's translation).

ferent system, then the signatories to the Treaty of Vienna must meet to sanction the changes—or to forbid them. . . . This time Bismarck, the man of moderation, snarled almost shockingly and showed his teeth. . . . he was the minister-president of a Prussia which had just revealed herself as the military superior of Austria and almost certainly of France as well. . . . now he showed that if Prussia could not get her own way (meaning his, Bismarck's, own way), he was ready to plunge all Europe into revolutionary violence.[38]

Between Prussia's victory at Königgrätz and the end of 1867, Bismarck implemented his *kleindeutsch* policy, leading quickly to the unification of northern Germany. The Treaty of Nikolsburg had ended Austrian hopes for a *grossdeutsch* policy that envisioned a unified Germany including both Prussia and Austria. Bismarck's *kleindeutsch* policy dominated Central European affairs until Hitler—an Austrian—launched the *Anschluss* that brought Austria back into greater Germany. By means of a war fought over lesser issues, Bismarck had unified much of Germany, prompted the formation of Austria-Hungary, unleashed nationalist sentiments in Eastern Europe, and brought the Congress of Vienna system to a crashing end. No wonder that contemporary observers were astounded by the war's development.

Events of so startling a character have taken place in the theater of war and present such an aspect of importance toward the future that the mind is dizzied by its attempt to estimate their real importance.[39]

Thirty dynasties have been swept away, . . . the fate of twenty millions of civilized men has been affected for ever, the political face of the world has changed as it used to change after a generation of war. . . . Prussia has leaped in a moment into the position of the first Power of Europe.[40]

The contemporaneous accounts reflect not only the flushed excitement of the moment, but also depict an enduring perspective on the consequences of the Seven Weeks' War. In this regard, it is worth contemplating Grützner's comparison of the Seven Weeks' War to the Franco-Prussian War. He argues that "the secular significance of 1866 has been underestimated by many German historians of the older generation. . . . Rather, the Franco-Prussian War would today be classified as only an epilogue to the policies inaugurated in 1866."[41] The Seven Weeks' War,

[38] Crankshaw (fn. 20), 219, 220–21.
[39] *Illustrated London News* 29 (July 14, 1866), quoted in Gordon A. Craig, *The Battle of Königgrätz: Prussia's Victory over Austria, 1866* (Philadelphia: Lippincott, 1964).
[40] *The Spectator* (London) 39 (July 7, 1866), quoted in Craig (fn. 39).
[41] Grützner (fn. 26), 80. "Die säkulare Bedeutung des Jahres 1866 unterschätzen viele deutsche Historiker der älteren Generation. . . . Dagegen wird heute der deutsche-französische Krieg eher als Epilog einer bereits 1866 inaugurierten Politik eingeordnet" (author's translation).

"der Bruderkrieg" as it is sometimes called, was motivated by Prussia's ambitions *within* the German community; it was expected to be of only minor consequence for the international status quo in Europe. In defeating Austria, however, Prussia overturned that status quo. How can we explain the emergence of German hegemony and the reordering of the post–Congress of Vienna international system as a consequence of a war fought over issues deemed by the theory of hegemonic war or of power transition to be too inconsequential to have such results?

War Costs and the International Interaction Game

The international interaction game provides a missing link in the puzzle of the Seven Weeks' War and the emergence of German hegemony. We have seen that war can be an equilibrium outcome even though the status quo is highly valued. But such a high valuation of the status quo imposes limitations on the expected magnitude of the ensuing war. If the stakes are expected to be relatively small—as in the case of the Seven Weeks' War—then the expected costs of the war *must* be low. If we assume that actual costs are correlated with observed costs, then wars between rivals who value the status quo highly are likely to be small wars.

Austria's demands were minor enough in terms of the overall international order; she sought to hold on to her first place among the German states. The Habsburgs did not (and were not expected to) demand territory from the Prussians. Nor did they demand a change in Prussia's external policies or any diminution in Prussia's role in the German Diet. They did not demand any fundamental change in Prussian policy; all they expected was continued deference to Austria as the first among equals. Likewise, Prussia's demands *before* the war were modest, although the possibilities opened by Bismarck proved not to be.

Prussia demanded a share in the title of first among equals in the German Confederation; she wanted to play a leading role in Germany, but did not demand an inch of Austrian soil. To be sure, Austria's acquiescence to Bismarck's demands would have increased Prussia's influence within the Diet over the disposition of Holstein. Still, the demands themselves were not enough to alter the perception of Prussian power in Europe fundamentally, as is suggested by the analysis of money market discount rates and by the review of Prussian policy before the war. This opinion is reinforced by the interpretation of several historians that the war, rather than the issues that produced the war, changed history.[42]

[42] See, for instance, Simon (fn. 18), 22-38; Crankshaw (fn. 20), 189–223; Taylor (fn. 22), 123–40; Grützner (fn. 26), 80–117; Michaelis (fn. 35).

Thus it was that the war which ultimately transformed the international system was fought over relatively minor issues in the *international* context—issues so insignificant that several authorities suggest war could have been averted by the *purchase* of Holstein from Austria, a purchase Bismarck was willing to make.

To say that the *international* status quo was reasonably well liked by the combatants and that the magnitude of the competing demands was fairly small is akin to saying that Prussia and Austria could satisfy the equilibrium conditions for war only if the expected costs were even smaller than the expected stakes. Thus, according to the international interaction game, the Seven Weeks' War, even though it ultimately transformed the international system, was likely to be a small war as far as costs are concerned.

In fact, in terms of duration, of total battle deaths, and of battle deaths per million of population, the Seven Weeks' War is unusually small in scope as multilateral wars go. Battle deaths per million population, for instance, represent only about one percent of all such deaths in multilateral European wars since 1816. This is a level approximately one-tenth of the statistically "expected" level of losses for such a war[43] even though the war directly involved three great powers and eight lesser states, with a fourth major actor (France) playing a crucial role on the sidelines. Indeed, with the auxiliary assumption that observed costs and expected costs are highly correlated, we see that in a variety of subtle ways the costs of war across a large cross-section of such events are consistent with the expectations derived from the international interaction game's equilibrium conditions for war.[44]

Conclusion

The relationship between expected costs, stakes, and the status quo derived from the international interaction game reveals a significant, though unintended, selection bias in some research on great-power wars. Some analysts mistakenly infer that incidents which yield major consequences must have been major events themselves. But great consequences may result from small causes. The conditions stipulated in theories of hegemonic war or of power transition are not essential for the

[43] The data for this assessment are drawn from Melvin Small and J. David Singer, *Resort to Arms: International and Civil Wars, 1816–1980* (Beverly Hills, CA: Sage Publications, 1982).

[44] For a more detailed analysis of the anticipated costs of conflict in light of the theoretical expectations derived from the game, see Bueno de Mesquita and Lalman (fn. 12), and Bruce Bueno de Mesquita, "Big Wars, Little Wars: Avoiding Selection Bias," *International Interactions* 16 (forthcoming).

occurrence of system-transforming wars. The perspective that views system-transforming wars as *necessarily* involving some combination of extended battles, many casualties, participation by many nations, and usually the exchange of territory conflates causes with consequences. Contrary to the expectations of these theories, we have a case of a system-transforming war in which the stakes were rather modest and in which—despite the many states involved and the masses of troops assembled at the decisive battle—casualties were relatively modest and the fighting lasted only a short time.

The game-theoretic analysis suggests that, by relaxing the assumptions that assign special significance to the satisfaction of rivals with the international status quo and that restrict the purview of researchers to large, costly wars, we can complement the explanatory potential of hegemonic war theories. The game itself is not designed only to account for war among the great powers, nor even only to account for war. Rather, it is a general scaffolding for understanding a broad array of international interactions. In that regard, it differs markedly from the theories of power transition and of hegemonic war; yet, it is a complement rather than a substitute for them. These theories, and attendant empirical assessments, have drawn our attention to the prospective importance of power parity and differential growth rates in influencing the courses and outcomes of great wars.[45] The game-theoretic perspective reveals the added potential for explaining great-power wars within the same theoretical context as lesser wars, thereby diminishing the tendency to use ex post facto information in analyses. Because hegemonic wars are generally known to be so only *after* the fact, the development of theory is critical in helping to point to the risk of such events ex ante. This, I believe, is a characteristic that the game adds to the power-transition theory's already helpful focus on differences in rates of growth and power.

APPENDIX

Six assumptions establish the range of preference orderings over outcomes within the international interaction game:[46]

[45] See, for instance, David Garnham, "Power Parity and Lethal International Violence," *Journal of Conflict Resolution* 20 (September 1976), 379–94; Erich Weede, "Overwhelming Preponderance as a Pacifying Condition among Contiguous Asian Dyads," *Journal of Conflict Resolution* 20 (September 1976), 395–412; Henk Houweling and Jan Siccama, "Power Transitions as a Cause of War," *Journal of Conflict Resolution* 32 (March 1988), 7–102; Woosang Kim, "Power, Alliance, and Major Wars, 1816–1975," *Journal of Conflict Resolution* 33 (June 1989), 255–73.

[46] For a full explanation and justification of these assumptions, see Bueno de Mesquita and Lalman (fn. 12).

A1: Given subgame perfect behavioral strategies, players choose the strategy with the greatest expected utility.

A2: The outcome of war is probabilistic, with P^i being player i's subjective probability of gaining its demand, with $A, B \epsilon i$.

A3: The initiator of force gains its demand with certainty only if the adversary chooses to capitulate rather than to retaliate. The capitulating state loses with certainty.

A4: All nations prefer to resolve their crises through negotiations rather than to reach the same resolution *after* bearing the heavy human, material, and political costs of war.

A5: Each outcome of the crisis subgame has potential benefits and/or costs associated with it. We decompose the costs into constituent parts such that $\alpha, \tau, \gamma, \phi > 0$; and $\tau > \alpha, \Gamma$. α is the cost in lost life and property associated with fighting *away* from one's home territory; τ is the cost in life and property of fighting at home as the *target* of an attack; γ is the cost in life and property from a first strike to which the attacked party *gives in*; and ϕ is the domestic political cost (separate from lost life and property) associated with using *force*.

A6: The utility from gaining one's demands exceeds the utility from keeping the status quo, which in turn exceeds the utility from losing by acceding to an adversary's demands: $G > Q > L > 0$.

POSTREVOLUTIONARY MOBILIZATION IN CHINA
The One-Child Policy Reconsidered

By TYRENE WHITE*

DURING the 1980s, the study of Chinese politics was dominated by the premise that the death of Mao and the rise of Deng Xiaoping marked the beginning of a new political epoch. Some scholars argued that the political course set in motion at the famous Third Plenum of the Eleventh Central Committee of December 1978 signaled a fundamental, even revolutionary break with the past; others saw a slower drift from a monistic system toward a more consultative authoritarianism.[1] Either way, the deradicalized, demobilizational thrust of the post-Mao reforms appeared to augur China's transition to a "postrevolutionary phase" of political normalization, regularized decision making and institutionalized party rule.[2] Against that background, China's "one-child-per-couple" family planning policy, introduced in 1979, appeared to run counter to the post-Mao political premise. With its radical goal, collectivist values, and mobilizational flavor, the one-child policy bore many of the characteristics of a Maoist-era mass mobilization campaign.

To account for this contrast, some scholars considered the one-child policy to be an isolated exception to the demobilizational trend; they argued that the specter of a looming population crisis that would under-

* An earlier version of this paper was presented at the 41st Annual Meeting of the Association for Asian Studies, Washington, DC, March 17-19, 1989. The research presented here was supported by the Swarthmore College Faculty Research Support Fund; the preparation of this manuscript was made possible by a sabbatical from Swarthmore College and a postdoctoral fellowship from the Fairbank Center for East Asian Research, Harvard University. Special thanks to Jean Oi, James Kurth, and Kenneth Sharpe for their comments on earlier drafts. Any errors or deficiencies that remain are mine alone.

[1] Lowell Dittmer, *China's Continuous Revolution: The Post-Liberation Epoch, 1949-1981* (Berkeley: University of California Press, 1987); Harry Harding, "Political Development in Post-Mao China," in A. Doak Barnett and Ralph N. Clough, eds., *Modernizing China: Post-Mao Reform and Development* (Boulder, CO: Westview Press, 1986), 13-37; Harry Harding, *China's Second Revolution: Reform after Mao* (Washington, DC: Brookings, 1987); Tang Tsou, "Back from the Brink of Revolutionary-Feudal Totalitarianism," in Victor Nee and David Mozingo, eds., *State and Society in Contemporary China* (Ithaca, NY, and London: Cornell University Press, 1983), 53-88.

[2] Richard Lowenthal, "The Post-Revolutionary Phase in China and Russia," *Studies in Comparative Communism* 16 (Autumn 1983), 191-201; Dittmer (fn. 1), 11.

World Politics 43 (October 1990), 53-76

cut the modernization goals of the new elite left no alternative to a strict birth control program.[3] Others, however, assumed that, all reformist statements and policies notwithstanding, the one-child campaign was revelatory—that is, evidence of a mobilization regime at its intrusive, penetrative worst.[4] Still others maintained that the enduring mobilizational impulse revealed by the one-child campaign was an artifact of the previous revolutionary epoch that would slowly disappear during the transition to institutionalized party rule.[5] What these views had in common was their implicit reference to an underlying model of Leninist regimes, one in which mobilization cannot outlive the transition to postrevolutionary communism. This model was most fully elaborated by Richard Lowenthal, who argued that, as modernization proceeds, the political attributes of the revolutionary phase—personalized, charismatic rule, revolutionary class struggle, and mass mobilization in pursuit of radical political and economic goals—inevitably give way to the postrevolutionary politics of oligarchic rule, regularized decision making, and party institutionalization.[6]

The Lowenthal model has drawn criticism on many counts,[7] but the hypothesis that postrevolutionary regimes eventually lose the capacity and desire to use mobilizational means has met no serious challenge. Kenneth Jowitt, for example, warned against interpreting the transition as a smooth and unilinear one. He argued that the difficult process of transition from revolutionary consolidation to postrevolutionary rule—or in his words, from the task of mobilization to the task of "inclusion"—frequently triggers mobilizational responses from threatened party elites. He also rejected the notion of a strict dichotomy between the mobilization and postmobilization phases, arguing for the existence of "*amalgam-regimes*" that represent "fragile combinations" of the two. Stressing that coexistence did not imply compatibility, however, he con-

[3] Martin King Whyte, "Who Hates Bureaucracy? A Chinese Puzzle," in Victor Nee and David Stark, eds., *Remaking the Economic Institutions of Socialism: China and Eastern Europe* (Stanford, CA: Stanford University Press, 1989), 233–54; Joyce K. Kallgren, "Politics, Welfare, and Change: The Single-Child Family in China," in Elizabeth J. Perry and Christine Wong, eds., *The Political Economy of Reform in Post-Mao China* (Council on East Asian Studies, Harvard University, 1985), 131–56.

[4] Steven W. Mosher, *Broken Earth: The Rural Chinese* (New York: Free Press, 1983).

[5] Harding (fn. 1, 1986).

[6] Richard Lowenthal, "Development vs. Utopia in Communist Politics," in Chalmers Johnson, ed., *Change in Communist Systems* (Stanford, CA: Stanford University Press, 1970), 33–116; Lowenthal (fn. 2).

[7] For a cogent review of these critiques, see David Stark and Victor Nee, "Toward an Institutional Analysis of State Socialism," in Nee and Stark (fn. 3), 1–31. See also Constance Squires Meany, "Is the Soviet Present in China's Future?" *World Politics* 39 (January 1987), 203–30; Kenneth Jowitt, "Inclusion and Mobilization in European Leninist Regimes," *World Politics* 28 (October 1975), 69–96.

curred with Lowenthal on the fundamental conflict between mobilizational and postmobilizational methods of political rule, describing the two methods as stark opposites. According to Jowitt, mobilization is characterized by the "controlled or elite-directed *disruption*" of "social, personal, institutional, and psychological" routines, as opposed to "the *methodical* consideration and management of tasks" characteristic of postrevolutionary rule.[8]

Do these propositions fit the Chinese case? The brutal crackdown on the 1989 democracy movement was a perfect illustration of Jowitt's "threat-reaction" model of elite response to a regime-threatening crisis. In this essay, however, I will argue that mobilization has remained an integral, active part of the postrevolutionary Chinese political process, not merely a reactive tool of fearful conservatives. Using the one-child campaign as a case study, I will demonstrate that the instrument of mobilization in Chinese politics, the mass campaign, was inherited by Deng Xiaoping as a part of the institutional legacy of the Maoist regime. Despite their unambiguous repudiation of Maoist campaign methods in 1978, the post-Mao reformers sought to harness them in practice for their own purposes, transforming an instrument of calculated disruption into an instrument of methodical, engineered change. As a result, the form, content, and intensity of the campaign changed over the course of the decade, but mobilization remained an essential instrument of communist party rule, particularly in the countryside.

Mobilization after Mao

One of the defining characteristics of the Maoist era in Chinese politics (1949–1976) was the continual use of "mass mobilization campaigns" (*qunzhong yundong*) to achieve socialist goals. Although mobilizational methods were a standard tool of political organization in other Leninist regimes, the Chinese campaigns were distinctive for the "utter seriousness" with which political study and ideological transformation were pursued.[9] For Mao, revolutionary goals were primary; under his leadership, China's revolutionary politics oscillated between two types of intensive mass participation and political struggle. The first was "storming," the distinctively Maoist model of mobilization from below, in which the

[8] Ibid., 93 and 94–95. Jowitt's challenge goes far toward acknowledging a continuing role for mobilization in postrevolutionary regimes, but whereas he views the continued use of mobilization to be *negative* or *reactive*, a "response to the problems and challenges" of expanding the bases of regime support, I view it as a *positive instrument* of power.

[9] Andrew Walder, *Communist Neo-Traditionalism: Work and Authority in Chinese Industry* (Berkeley: University of California Press, 1986), 121.

elite party membership is subjected to rectification by the nonelite masses. The second was "engineering," the Maoist-Liuist-Leninist model of party-controlled mobilization from above in pursuit of specific and practical goals.[10]

Campaigns served at least three vital functions during the Maoist era. First, they helped the party transcend its own organizational limits by recruiting and leading mass participants in a wide variety of tasks—everything from implementing new marriage laws, to disease control, to eliminating counterrevolutionaries.[11] This function was particularly important during the 1950s, when mass involvement was needed to reinforce the party's administrative capacities and to indoctrinate and mobilize a newly liberated population. Second, campaigns were tools for correcting various bureaucratic and political ills. They were used to rectify the party, promote discipline and unity, purify the state bureaucracy, and prevent the entrenchment of elite-mass cleavages. Third, campaigns were used as populist instruments of permanent revolution to "smash" the party/state apparatus.[12]

The post-Mao leadership formally repudiated all three functions of the mass campaign. Mao's culminating effort at smashing the party/state during the Cultural Revolution had such force and consequence that the post-Mao elite, victims of the process, formally abandoned mobilizational politics and fostered the process of institution building.[13] The new leaders replaced mass campaigns and permanent revolution with routinization and modernization as the primary means and ends of state power in post-Mao China. Radicalism gave way to "systems engineering," with all its attendant overtones of incremental manipulation of countless social, political, and economic levers. And populist, revolutionary methods of mass participation gave way to imperfect electoral mechanisms, expanded public debates and channels for political expression and influence, glorification of the rule of law, and a shrinking monopoly of party power. In short, the post-Mao elite fundamentally redefined its historical mission, abandoning the rapid achievement of revolutionary ends by means of class struggle in favor of the steady development of the "primary stage of socialism" and the construction of a "socialist spiritual civ-

[10] Dittmer (fn. 1), 6. For other discussions of mobilization, see Squires Meany (fn. 7), and Walder (fn. 9), 113–22.

[11] Charles P. Cell, *Revolution at Work: Mobilization Campaigns in China* (New York: Academic Press, 1977), 106–7.

[12] Harry Harding, *Organizing China: The Problem of Bureaucracy, 1949–1976* (Stanford: Stanford University Press, 1981), 331–33; Dittmer (fn. 1), 6–8.

[13] Squires Meany (fn. 7); Tang Tsou (fn. 1).

lization."[14] In so doing, the elite brought the Maoist revolutionary phase to an end.

The question is whether the end of the revolutionary phase meant the end of party-led mobilization, as was the case in the Soviet Union.[15] Despite the formal repudiation of mass movements, only the third type of mobilization campaign disappeared entirely in the 1980s. Mobilization was still employed to achieve political goals, such as party rectification (1983–1986), eliminating spiritual pollution (1983), and countering bourgeois liberalization (1987 and 1989). It was also used to achieve substantive goals, such as obligatory tree planting, electoral reform, and controlling economic crime.[16] Nowhere has the use of mobilization been more sustained or controversial, however, than in the arena of family planning.

A. The "One-Child-per-Couple" Mobilization Campaign, 1979–1983

The effort to limit couples to only one child began just as China's new leadership was repudiating Maoist campaign methods. There is no shared, unambiguous definition of a mass campaign or its precise form, but the family planning campaign implemented between 1979 and 1983 fulfilled and exceeded every basic requirement. Two studies of campaigns and their dynamics, Gordon Bennett's *Yundong* and Charles Cell's *Revolution at Work*, provide working definitions. Bennett defined the campaign as "a government-sponsored effort to storm and eventually overwhelm strong but vulnerable barriers to the progress of socialism through intensive mass mobilization of active personal commitment."[17] Cell relied on the definition provided by a 1959 article in *Hongqi* (Red Flag), which described the campaign as an "organized mobilization of collective action aimed at transforming thought patterns, class/power relationships and/or economic institutions and productivity." More simply, he noted that "all campaigns involve an increased intensity of activity

[14] Zhao Ziyang, "Advance Along the Road of Socialism with Chinese Characteristics," *Beijing Review* 45, November 9–15, 1987, pp. I–XXVII.

[15] In the Chinese case, the communiqué of the Third Plenum of the Eleventh Central Committee in December 1978 and the resolution on CCP history adopted in June 1981 marked the end of the revolutionary phase. The parallel moment in Soviet history was the party program adopted at the 22nd Party Congress in October 1961. On the Soviet case, see Lowenthal (fn. 2).

[16] Lester Ross, "Obligatory Tree Planting: The Role of Campaigns in Policy Implementation in Post-Mao China," and Barrett L. McCormick, "Leninist Implementation: The Election Campaign," both in David M. Lampton, ed., *Policy Implementation in Post-Mao China* (Berkeley: University of California Press, 1987), 225–52 and 383–413.

[17] Gordon A. Bennett, *Yundong: Mass Campaigns in Chinese Communist Leadership* (Berkeley, CA: Center for Chinese Studies, 1976), 18.

beyond what is expected in regular work and living routines"; "wha may not start out as a campaign may in due course become one, or vic versa."[18]

Both definitions suggest two irreducible characteristics of the mas. campaign. First, its substantive goal is to advance socialism by targeting a particular obstacle and/or by promoting a particular ideal. Second, the procedural requirements for a mass campaign are that it be organized and launched from above, led by the party and mass organizations, and that the masses be mobilized beyond their normal routines. The one-child campaign met all of these criteria.

First, with the immediate task of building socialism redefined to be the achievement of the four modernizations, promoting the policy of one child per couple not only explicitly aided economic goals; it also targeted old habits and customs that had been variously described as decadent or feudal, and that stood as barriers to socialist development. For example, such traditional sayings as "men are superior to women" and "the more children, the more happiness" became objects of condemnation in propaganda campaigns, as did the associated behavior of female infanticide.[19] The policy of promoting only one child per couple and denouncing the traditional preference for sons was tantamount to launching a narrowly defined cultural revolution.

Second, beginning in 1979, the call for the one-child family led to nationwide mobilization. Hua Guofeng set forth the impossible goal of reducing the population growth rate to five per thousand by 1985; Chen Muhua, head of the State Council Leading Group for Family Planning, called for a mass campaign to achieve this target.[20] Provincial leaders responded by holding mobilization meetings attended by representatives of all party, government, and mass organizations; recruiting family planning activists and propagandists from various units; sending medical and propaganda teams into the countryside to promote birth control; and publishing articles glorifying the one-child family.[21]

Subsequently, implementation of the one-child program involved the use of standard campaign methods. Both Bennett and Cell listed specific

[18] Cell (fn. 11), 7.

[19] *Foreign Broadcast Information Service, Daily Report—China* [hereafter cited as *FBIS*], February 15, 1980, p. L14; *Renmin Ribao*, September 25, 1980, p. 1.

[20] *FBIS* (Supplement No. 15), July 2, 1979, p. 19; Sichuan Provincial Service, December 17, 1979, in *FBIS*, December 18, 1979, p. Q3; Li Xiuzhen, "Dangqian jihua shengyu gongzuode xingshi he renwu" [The current situation and tasks in family planning work], *Renkou Yanjiu* 1 (1980), 3.

[21] Tyrene White, "Population Policy and Rural Reform in China, 1977–1984: Policy Implementation and Interdependency at the Local Level" (Ph.D. diss., Ohio State University, 1985), chap. 3.

indicators of campaign activity, with Cell providing the most exhaustive list.[22] Three types of indicators emerge from his classification: first, *informational* indicators (newspaper articles, slogans, mobilization meetings, targets for criticism, stories for emulation, pamphlets, signs, banners, exhibits); second, *organizational* indicators (sending in outside cadres, creating work teams, reallocation of resources, curtailment of other programs or activities, disruption of work routines or reorganization of the unit); third, *mass participation* indicators (letter writing; participating after regular work hours; mobilization of minority, youth, or other special groups and organizations; study groups; struggle and/or criticism sessions; local, regional and national rallies).

Virtually all of the above tools were used between 1979 and 1984 to implement the one-child policy.[23] Telephone conferences and mobilization meetings were employed to energize cadres and to disseminate campaign targets. Special medical teams were sent to perform sterilizations and abortions, or to insert intrauterine devices. Propagandists filled newspapers with model stories and educational editorials, and cadres who violated the new policy were used as negative models for public discussion. In rural areas, where the greatest resistance was encountered, local party cadres held as many as nine or ten mobilization meetings per year in one locality.[24] Normative and material incentives, such as glorification meetings and cash rewards, were used to encourage compliance. Those who resisted, however, encountered coercive measures. In the most extreme cases, women were forcibly taken to clinics for abortions or other procedures; more typical, however, was the practice of making women attend endless meetings without pay, or subjecting couples to propaganda and harassment in their homes.[25]

Although the struggle sessions against offenders did not take on the intensity of those against political targets during the Maoist period, individuals, and wayward cadres in particular, were isolated and criticized publicly as examples. Their punishment frequently went beyond eco-

[22] Bennett (fn. 17), 38–74; Cell (fn. 11), 92–110.

[23] The primary literature on this period is vast. For analyses and descriptions, see Tyrene White, "Implementing the 'One-Child-per-Couple' Population Program in Rural China: National Goals and Local Politics," in Lampton, ed. (fn. 16), 284–317; Mosher (fn. 4), 224–[illegible]; John S. Aird, "Coercion in Family Planning: Causes, Methods and Consequences," in U.S. Congress, Joint Economic Committee, *China's Economy Looks toward the Year 2000*, vol. I: *The Four Modernizations* (Washington, DC: Government Printing Office, 1986), 184–221.

[24] Yuan Fang and Gu Shengzu, "Nongye shengchan zerenzhi qianhoude jihua shengyu guanli tizhide bijiao fenxi" [A comparative analysis of the managerial system of family planning work before and after the agricultural responsibility system], *Shehuixue yu shehui diaocha* ([illegible]1987), 7–16.

[25] On these practices, see Aird (fn. 23) and Mosher (fn. 4).

nomic sanctions, even to the point of obligatory sterilization.[26] Conversely, cadre mobilization and emulation efforts were taken very seriously. In one Hubei village, the local "women's leader" (a female cadre in charge of women's affairs and family planning in each of China's production brigades and teams) attributed her brigade's exemplary record on birth control to the example set by the local cadres. She and the brigade's eight team-level women's leaders had undergone "voluntary" sterilizations when the one-child campaign began.[27] Thus did they demonstrate their willingness to "take the lead" in family planning and set a good example for the masses.

In short, though the scope of the family planning campaign did not rival the radical political campaigns of the Maoist era, it was a classic mobilization drive of major proportions. Over time, however, pressures to bring campaign tactics in line with the requirements of an ongoing process of rural reform gave rise to a variant form of mobilization politics.

B. The Politics of Institutionalized Mobilization, 1984–1989

Although mobilization for family planning was continuous in the early 1980s, the environment for implementing the one-child policy had changed dramatically by the end of the Sixth Five-Year Plan in 1985. At the time the campaign was launched in 1979, the regime of Hua Guofeng was fostering a mobilizational style of politics that had direct antecedents in the Maoist era. By the end of that year, however, political campaigns had been rejected as tools of policy implementation, and the reform coalition of Deng Xiaoping was rapidly consolidating power. As a result, the traditional language of mass mobilization was softened even while the demand for a one-child norm became more insistent.

By September 1980, cadres were exhorted not to rely on coercion to enforce the policy; they were told to make family planning a part of their routine work. The one-child campaign continued, however, with the Central Committee and the Communist Youth League issuing an unprecedented "Open Letter" to all party and youth league members urging them to "take the lead" in having only one child.[28] In the cities, effective enforcement was made possible by the close surveillance of

[26] Beijing City Service, April 12, 1980, in *FBIS*, April 21, 1980, p. R1; *Renmin Ribao*, April 11, 1980, p. 4; *Kyodo*, June 26, 1980, in *FBIS*, June 27, 1980, p. Q1.

[27] Interview file, by the author, May 23, 1982, Huashan commune. This interview was conducted as part of two periods of rural fieldwork in China (February–December 1982 and September 1983–June 1984). The commune (now reclassified a rural town) is situated to the northeast of Wuhan municipality, Hubei Province. Additional information on the field site and interview procedures may be found in White (fn. 21).

[28] *Renmin Ribao*, September 25, 1980, p. 1.

women in the workplace and in neighborhoods, but rural enforcement posed enormous obstacles. Not only did the peasants strongly resist the one-child limit, but the pattern of rural organization underwent radical changes between 1980 and 1983 as the defining institutions of the Maoist era were dismantled. Household farming on collectively owned land replaced collectivized agriculture, separate organs of township government and economic management supplanted the people's commune, and the revival of free markets broke the party's absolute grip on rural income.

These structural changes diminished the political and economic power of rural cadres and substantially reduced the dependency and vulnerability of the peasantry. The changes also weakened a vital component of the local enforcement structure for family planning—grassroots women's leaders in production brigades and teams.[29] As local activists, these women had provided the crucial link between the party hierarchy and rural women of child-bearing age by monitoring pregnancies and "mobilizing" for compliance; they now found that the political rewards of cadre service did not outweigh the material rewards of household production. Like other laborers, they were allotted plots of land and were obliged to fulfill their household responsibility contracts. At the same time, pressures to reduce the number of "nonproductive" cadres deprived them of the small administrative subsidy once provided by the brigade. Finally, in the altered political climate of the 1980s, women's leaders found themselves not just deprived of the status once implied in cadre service, but in danger of assault or abuse for their role in the family planning campaign.[30]

Although the situation stabilized by 1984, these changes provoked a sense of panic about family planning work during 1982 and 1983, as officials feared that grassroots organizational disarray or paralysis was undercutting the one-child campaign. To compensate for rising birth rates and organizational deficiencies, a sterilization campaign had been launched in late 1982 that continued sporadically through 1983 and 1984. It was specifically directed at couples who were under the age of forty and already had two or more children, and it was designed to reduce the incidence of third or additional births.[31] The full dimensions of the mo-

[29] White (fn. 23).

[30] One report from a county in Jiangsu Province indicated that there had been 381 "incidents of revenge" in twelve towns and townships in an eighteen-month period, of which 32 percent were related to family planning. See Su Suining, "There Are Many Causes of Strained Relations between Cadres and Masses in the Rural Areas," *Nongmin Ribao*, September 26, 1988, p. 1, in *FBIS*, October 7, 1988, p. 13.

[31] *Xinhua*, November 1, 1982, in *FBIS*, November 3, 1982, p. K20; Qian Xinzhong, "Nuli kaichuang jihua shengyu xuanchuan jiaoyu gongzuo xin jumian" [Make great efforts to initiate the new phase of family planning propaganda and education work], *Renkouxue Kan*

bilization remain unclear, but by the end of February 1983, nearly nine million "birth control operations" had been performed nationally, a figure 1.6 times the total for all of 1982.[32] In Hubei Province alone, one million sterilizations occurred over the course of 1983.[33]

By the end of 1983, other elite priorities began to overtake the one-child campaign. A party rectification campaign was launched, a central feature of which was to correct the bad "work style" of many party members and to improve party-mass relations. In addition, the reformers were poised to push through new rural reforms at the end of 1983 that would commercialize the rural economy and break up the party's monopoly of supply and distribution. At the same time, the township governments that had replaced people's communes were placed on a tighter fiscal leash, with expenditures for basic activities like family planning set in an annual fixed budget. In this shifting political, economic, and institutional context, a new central document on family planning was issued in 1984. Central Document 7 departed from its predecessors by allowing some restricted exceptions to the one-child limit, exceptions that applied to about 5 percent of all couples of child-bearing age among the majority Han population (rules for minorities were slightly less strict). More importantly, the document repudiated the use of coercion to enforce the one-child limit, stressing instead the need for routine and patient work in order to gain compliance.[34] Subsequently, family planning was declared to be a fundamental state policy: the government stressed the obligation of Chinese citizens to engage in family planning, and made efforts to build an organizational structure for routine administration. Yet the gap between official family planning targets and the capacity of the formal party/state apparatus to reach them undermined efforts to regularize grassroots implementation. Instead, the implementation process settled unstably into a pattern of *institutionalized mobilization*.

Institutionalized mobilization is a variant form of implementation, which involves periodic, functionally defined mobilization efforts that (1) temporarily intensify coercive and normative incentives; (2) vary from region to region in timing, intensity, and scope; (3) last for limited, predictable periods of time (often specified at the outset); (4) have as their

1 (1983), 9; *Xinhua*, January 10, 1983, in *FBIS*, January 13, 1983, pp. K13–14; Beijing Domestic Service, January 27, 1983, in *FBIS*, March 3, 1983, p. K11; Shi Chengli, "Wo guo jihua shengyu gongzuode fenqi" [Analysis of our country's family planning work], *Xibei Renkou* 1 (1988), 31; interview conducted by the author, June 8, 1984, Family Planning Commission, Hubei Province.

[32] Shi (fn. 31), 31.

[33] Interview (fn. 31).

[34] *Jihua Shengyu Ban, Jiankang Bao*, March 9, 1984.

primary goal behavioral control or "practical results," not attitudinal or cultural change; (5) have a diminished scope of mass participation in favor of narrow mobilizations of the target population; and (6) utilize extensive propaganda to shape public sentiment, but discourage disruptive mobilizational activities beyond the target population in order to insulate the project from economic production and other reform initiatives.

Several aspects of the family planning campaign illustrate the pattern of institutionalized mobilization. The family planning bureaucracy and delivery network has been strengthened, funding and personnel have increased, routine procedures have been developed to administer the work and coordinate with other relevant departments (for example, public health and civil affairs), and supplemental regulations have been drafted to clarify policy. Family planning work has been thoroughly professionalized; demographers, sociologists, and other social scientists are involved in the policy-making and implementation processes, and full-time family planning workers are trained and placed at the local level. National meetings are held regularly in order to assess ongoing work. Grassroots cadres have forms, procedures, and timetables for gathering data on local demographic trends. New loopholes in the control system, such as the problem of unmonitored peasants who migrate temporarily to the cities, are closed as new procedures are developed—in this case by requiring local commercial bureaus and offices to withhold enterprise licenses from individuals who violate family planning regulations.[35]

None of these developments has obviated the need for mobilizational methods, however. The strict one-child limit has been progressively relaxed since 1984 in order to make the policy "acceptable to the peasants."[36] As a result, rural couples whose first-born is a daughter are automatically eligible for a second child.[37] This relaxation has partially accommodated the demand for male offspring that frustrated earlier enforcement efforts, but it has not put an end to peasant resistance. Couples with one son or two daughters often bear additional children; the required four-year interval between the first and second birth is routinely ignored; and village leaders conspire with relatives and friends to conceal "excess" births from higher authorities.

[35] De Ming, "China's Population Situation Remains Grim," *Liaowang Overseas Edition* 17 (April 25, 1988), 10–11, in *FBIS*, May 11, 1988, p. 28.

[36] *China Daily*, April 19, 1988, p. 1.

[37] *Zhejiang Ribao*, September 27, 1987, in *FBIS*, October 13, 1987, pp. 31–32; *China Daily*, July 30, 1987, p. 3; Zhao Yugui, "Zhongshi liudong renkoude jihua shengyu guanli" [Pay attention to the management of family planning work for the floating population], *Renkou yu Jingji* 2 (1988), 5.

Continuing resistance to a one- or two-child limit is not the only motive for mobilization. Because the targeted age cohort is constantly replenishing itself, because today's compliant couples may become tomorrow's problem, and because overworked rural leaders find it convenient to neglect family planning, only periodic mobilization can indoctrinate young couples and emphasize the determination of the state to prevail. Family planning propaganda and education campaigns occur predictably in the summer and winter; the campaigns are coordinated with agricultural slack seasons, the corresponding national work conferences that review local and national statistics, and the Chinese New Year (a time for spousal reunions and marriages). They still require the temporary mobilization of propagandists, family planning workers, and medical personnel for grassroots work. Propaganda activity and intensive indoctrination have ceased to be the centerpiece of the campaign, however. Instead, the focus is on the achievement of practical results by carrying out medical procedures such as sterilizations, abortions, and IUD insertions on the spot.[38] For example, in the fall of 1987 Zhejiang Province launched a "family planning publicity and technical service campaign," in which "areas lagging behind" in family planning work were urged to "strive for a marked improvement during the campaign."[39] Similarly, a circular prepared for family planning month in Yunnan (January 1987) stressed the need to "mobilize the masses to vigorously and conscientiously implement the measures for birth control and remedial measures" (that is, abortions).[40]

Although national officials like Peng Peiyun, head of the State Family Planning Commission, stressed that cadres "should not neglect regular work and then merely rely on temporary mobilization work,"[41] local enforcement still revolves around centrally disseminated targets and quotas that trigger periodic mobilizations and "shock drives" at the local level. A classic example of this pattern is the May 1985 campaign that was waged in a suburban village outside Xiamen municipality, Fujian Province, and reported by Huang Shu-min.[42] In May 1985, the village leaders suddenly announced a strict new policy on having a second child and ordered all women of child-bearing age to report for examinations to detect pregnancies. The decision was made summarily and backed up

[38] *Zhejiang Ribao*, July 25, 1986, p. 1.

[39] *Zhejiang Ribao*, September 27, 1987, in *FBIS*, October 13, 1987, 31–32.

[40] Yunnan Provincial Service, January 5, 1987, in *FBIS*, January 7, 1987, p. Q2.

[41] Yang Chaoling and Li Qin, "Peng Peiyun, Minister of the State Family Planning Commission, on China's Family Planning Situation and Policy," *Liaowang Overseas Edition* 17 (April 1988), 9–10, in *FBIS*, May 2, 1988, p. 35.

[42] Huang Shu-min, *The Spiral Road: Change in a Chinese Communist Village through the Eyes of a Communist Party Leader* (Boulder, CO: Westview Press, 1989), 175–85.

by the terse announcement that those who failed to comply would be cut off from village business activities. Despite the reforms, this threat was extremely potent, since village approval for permits and loans smoothed the way for many of the private economic activities that produced the bulk of income for many village households. No educational meetings were held, no propaganda announcements were made. On the appointed day, however, a special work team from the township arrived in the village to consult with the village party secretary and to transport pregnant women to the township hospital. The work team was one of several sent out by the township, each assigned to cover a group of villages.

Although the events in the village were straightforward, they do not reveal how the mobilization process made its way to the village level or explain the timing of the campaign. Media reports from the period reveal that, in standard campaign fashion, the impetus came from higher levels of government. During 1984 and 1985, most provinces responded to new central guidelines on family planning by adopting new or modified regulations at meetings of their provincial people's congresses. Because 1985 was the final year of the Sixth Five-Year Plan, extra pressure was placed on all provinces and localities to meet the targets of the plan. The approach taken by each province varied, however; the timing tended to coincide with the convening of the provincial people's congress, the official legislative body of the province. In Fujian, the Third Session of the Sixth People's Congress was held in late April and early May, at which time a special forum on family planning was held. According to a radio report on the meeting, the delegates were told that they "must do a good job on work in May and June, a key period on which the success of the population control tasks in the whole year depends."[43] Once this provincial directive was translated into specific quotas for each county and township, it resulted in the deliberate and methodical village-level mobilization.

In 1987, Sichuan provincial authorities launched a summer "propaganda, inspection and implementation drive";[44] in Hainan, regional officials organized a springtime "shock drive." Cadres were instructed to

> summon up the greatest resolve to organize the family planning shock drive, and achieve unified leadership, arrangements, and action. . . . It is necessary to focus on remedial measures regarding pregnancies not covered by the plan, and also do a good job in carrying out ligation surgery after the birth of a second child and inserting an intrauterine device after the birth of a first.[45]

[43] Fujian Provincial Service, May 3, 1985, in *FBIS*, May 9, 1985, p. O1.
[44] Sichuan Provincial Service, June 11, 1987, in *FBIS*, June 12, 1987, p. Q1.
[45] Hainan Island Service, May 1, 1987, in *FBIS*, May 4, 1987, p. P1.

Two weeks later, regional officials reported that, between April 27 and May 9, more than 8,700 procedures, including 4,244 tubal ligations, had been performed.[46] After criticizing areas that were still lagging behind, they called for a "new high tide," and issued specific instructions:

> Prior to the end of June, all cities and counties must accomplish over 50 percent of the special annual family planning tasks set by the administrative regional authorities.
>
> Second, in bringing about the high tide . . . all cities and counties must regard the task of collecting fees imposed on those who have exceeded the limit to the number of births as an important task, and ensure the implementation of measures for sterilization through the promotion of the task of collecting fees imposed on those who have exceeded the limit to the number of births.
>
> Third, public health departments . . . must send a number of medical personnel with high political and professional standards to serve people undergoing sterilization operations.[47]

When Hainan officials met in February 1988 to review the previous year's work and plan for the future, their report captured precisely the effort to fuse regular work with campaign tactics:

> We must establish and put on a sound basis family planning laws, rules and regulations, and beef up the family planning organs and personnel. Shock family planning drives must be regularly organized and combined in an organic way, with the emphasis on practical results.[48]

Weak grassroots organization also makes mobilization a necessity. Professionals and full-time family planning workers may staff family planning organs at or above the county level, and sometimes at the township (previously the commune) level; below the township level, family planning, if done at all, is conducted by party cadres and women's leaders—that is, local political leaders and activists.[49] After 1984, when structural reform of the communes was completed, all township governments were supposed to be staffed by a full-time cadre for public health and family planning. In many areas, however, townships have refused to pay such individuals out of collective revenues, but the state has not put them on its payroll either. As a result, many townships have no family planning cadre, or one that is poorly paid.[50] At the village (brigade) and small

[46] Ibid.; Hainan Island Service, May 18, 1987, in *FBIS*, May 20, 1987, p. P2.

[47] Ibid., pp. P2–3.

[48] Hainan Island Service, February 17, 1988, in *FBIS*, February 18, 1988, p. 21.

[49] *Henan Nianjian, 1985* [Yearbook of Henan Province, 1985] (1986), 597. By 1987, Qinghai Province reported that 30% of all townships and towns had no family planning organs. Qinghai Provincial Service, February 20, 1987, in *FBIS*, February 27, 1987, p. T2. See also Liaoning Provincial Service, February 28, 1988, in *FBIS*, March 3, 1988, p. 38.

[50] Qinghai Provincial Service, February 20, 1987, in *FBIS*, February 27, 1987, p. T2. See

group (team) levels, family planning came under the purview of the local women's leader, but team-level cadres slowly disappeared after 1984 and village leaders often refused to employ a family planning cadre, leaving the job of enforcing local birth quotas to unpaid representatives of the Women's Association. Even where cadres are in place, however, many of them are villagers first and cadres second where family planning is concerned. They will respond to signals from higher authorities, but professionalization does not reach into the village. In short, functional specialization deteriorates rapidly at the township level and below, and neither the collective nor the state has the funds to remedy the problem entirely. Instead, family planning associations have been organized down to the village level in the hope that locally mobilized and functionally defined mass organizations will contribute to the stabilization of the work.[51] Thus, at the grassroots, politically weak, poorly skilled, and underpaid women's leaders, activists, and association members rely on local party mobilization in order to cope with recalcitrant villagers. In 1988, this dependency on the party apparatus led central leaders specifically to exempt family planning from ongoing efforts to reduce party interference in routine government work.[52]

Finally, although the intensity of family planning propaganda has declined since 1985, family planning work has constantly been linked to other political, ideological, or morality movements, including party rectification, the antibourgeois liberalization campaign, and the call to build a socialist spiritual civilization. The linkage is designed to reinforce the collective value and larger moral purpose of individual decisions to comply voluntarily with family planning goals, and to associate policy violators with politically suspect, morally bankrupt, or criminally inclined elements.

In sum, implementation of the one-child policy continues to meet the two primary characteristics of a mobilization campaign—organized, routine-altering activity for the purpose of defeating a barrier to socialist development. Since 1984, however, the predictability, decentralization, and routinization of the mobilization process has transformed the campaign into a mobilizational instrument of some precision. While the Communist Party is unwilling to alienate the critical pro-reform peasant

also Liaoning Provincial Service, February 28, 1988, in *FBIS*, March 3, 1988, p. 38. In some cases, "family planning cadres" were individuals who held more than one work portfolio simultaneously. See Shuai Zepeng, "Jihua shengyu ganbu duiwude guanli jidai jiaqiang" [Management of family planning cadre ranks in urgent need of strengthening], *Renkou Yanjiu* 4 (July 1988), 59. *Tianjin Ribao*, March 13, 1990, in *FBIS*, April 11, 1990, pp. 45–46.

[51] *FBIS*, February 19, 1988, p. 23.

[52] *Xinhua*, January 20, 1988, in *FBIS*, January 21, 1988, p. 6.

*constituency by demanding strict enforcement of a one-child limit, it equally unwilling to channel more state funds into the development of rural family planning bureaucracy or to let resistance to birth contrc undermine its economic goals. The routine bureaucratic process ha therefore been joined to the party's grassroots mobilizational tradition tc yield a specific form of postrevolutionary mobilization. This hybrid is not simply a halfway measure between the artificially constructed polarities of revolutionary mobilization and rational-legal forms of enforce*ment, however. Instead, it is a new variation on the single political theme of Communist Party rule in China. The context for campaign implementation has changed, but mobilizational impulses and the institutional arrangements that produce them have endured.

C. The Limits of Institutionalized Mobilization

By early 1990, it was clear that China's family planning program was in jeopardy. The State Statistical Bureau reported steady population growth rates of about 1.4 percent between 1987 and 1989—rates higher than planned, but within an acceptable range. Surveys conducted under the auspices of the State Family Planning Commission, however, were showing growth rates as high as 1.6 percent; by January 1990 Peng Peiyun, the head of the commission, confirmed the existence of a gap of 30 percent between the official figures and the estimates of the commission.[53] Clearly, the state's efforts to achieve an average fertility level of about 1.7 children per couple were failing.

At first glance, this failure may appear to confirm Lowenthal's thesis. One might argue that the general trend in China is toward demobilization, and that the lingering presence of the one-child campaign reflects nothing more than the extraordinary nature of China's population dilemma. The regime's inability to sustain effective mobilizational efforts for more than a few years after the postrevolutionary transition merely confirms the long-term causal relationship between the modernization process and demobilization.

The flaw in this argument is that the failed effort cannot be equated with the end of mobilizational politics. The overall success or failure of a campaign never provided grounds for bringing the mobilizational character of the Maoist regime into question. Mao's two most famous efforts, the Great Leap Forward and the Cultural Revolution, were both spectacular failures if judged by his own initial goals. The Great Leap

[53] Qi Bangmi, "The Alarm Bell of Population Tolls Again," *Ban Yue Tan* 23 (December 1988), 15–16, in *FBIS*, January 18, 1989, p. 25; *Xinhua*, December 13, 1989, in *FBIS*, January 5, 1989, p. 5.

did not boost agricultural yields or accelerate China's march to communism, and the Cultural Revolution did not prevent the routinization of party rule. Yet the mass campaign was a defining characteristic of the Maoist period. In the post-Mao era, rising birth rates alone do not imply that campaigns are no longer employed. Campaigns continue to be the most effective means by which party leaders concentrate effort on specific tasks and overcome bureaucratic inertia and grassroots resistance. In addition, they fill an important gap in the state's capacities below the county level by prompting township- and village-level cadres to behave in ways that reflect the priorities of their superiors. By setting the goals and the time frame for achieving specific tasks, and by linking cadre evaluations to the fulfillment of those tasks, the party can increase the probability of effective grassroots implementation.

Several other factors provide a better explanation for the failure of institutionalized mobilization to achieve the ambitious goals of the central elite. To begin with, the goals themselves remain extremely ambitious, despite relaxations in the one-child rule. According to current policy, all urban couples are limited to one child, with very few exceptions. In the countryside, all couples whose first child is a girl are now eligible to have a second child after an interval of about four years; the same goes for couples who fit a variety of other categories (for example, those whose first child is seriously handicapped, or a husband who breaks tradition by moving into his wife's family home). No one is eligible for a third child, with the exception of certain small nationality populations living in sparsely populated areas. These guidelines may appear lenient by the impossibly strict rules of the early 1980s, but in China's agricultural society they are considered extremely strict and are difficult to enforce, even with campaign methods.

Moreover, for mobilization to be successfully deployed, at least one of three prerequisites must be met. The first concerns incentives. In earlier efforts to engineer change in the countryside, success depended on being able to offer material as well as normative incentives to a mobilizable constituency.[54] Great efforts were made to do just that in the early 1980s, when one-child couples received preferential health care, infant and child subsidies, extra plots of land, and other material rewards.[55] By 1984, however, as the one-child policy settled into a fundamental state policy, the rationale for special rewards for compliance began to erode, as did

[54] Dittmer (fn. 1), 39–43 and 237–38; Vivienne Shue, *Peasant China in Transition: The Dynamics of Development toward Socialism, 1949–1956* (Berkeley: University of California Press, 1980), 326–27.

[55] White (fn. 23), 284–317.

the local resources to provide them. In addition, pressures to reduce the expenses of township government led to cuts in family planning subsidies and bonuses.[56] As a result, by the mid-1980s cadres and bureaucrats urged family planning on grounds of delayed material gratification, a line that did little to offset preferences for multiple sons as the path to short-term happiness and long-term wealth and security.

Second, sustained and successful mobilization requires organizational stability.[57] Where turnover among local leaders or a target population is high, patterns of mobilization, study, and enforcement are constantly disrupted by personnel change. Although China's rural leadership has remained relatively stable, organizational changes, retirements, entrepreneurial opportunities, and elections have all contributed to a changing composition of the grassroots leadership. More importantly, however, the rural population—for the first time since the 1950s—now enjoys opportunities for mobility that complicate efforts to enforce family planning through either routine work or mobilization. The reported childbearing explosion among the "floating population" of fifty million has become the biggest headache for family planning officials, who refer to this group as an "excess birth guerrilla corps" and "excess birth floating villages."[58] The proposed solution—demanding that the urban commerce bureaus verify compliance with family planning before issuing local commercial licenses to migrant peasants—is no help, since commerce bureau officials have neither the time, inclination, or incentive to become adjunct family planning officials.[59]

Third, and most importantly, in the absence of compelling material rewards and organizational stability, mobilization requires a disciplined organization for enforcement—precisely the area in which communist parties traditionally excel. In this case, however, many cadres (native villagers themselves) have personally opposed and violated the one-child policy from the start. Beijing's disavowal of coercion and rigidly uniform implementation has simply given them greater leeway to interpret guidelines in ways that suit local preferences. Moreover, despite periodic nation-wide or localized mobilizations, the relatively lax political environment and stress on economic performance led many to report false statistics for family planning work in order to increase bonuses, avoid

[56] *Renmin Ribao*, June 12, 1984, carries the original text of Central Document 1 (1984). For a translation, see *China Quarterly* 101 (March 1985), 132–42.

[57] Walder (fn. 9), 114–20.

[58] Guo Xiao, "The 'Population Explosion' Is Drawing Near," *Jingji Ribao*, January 10, 1989, in *FBIS*, February 3, 1989, p. 51.

[59] *China Daily*, July 30, 1987, p. 3; Yunnan Provincial Service, January 3, 1988, in *FBIS*, January 5, 1988, p. 44; *Xinhua*, January 26, 1988, in *FBIS*, January 28, 1988, pp. 16–17; *Renmin Ribao*, June 4, 1985, p. 3.

sanctions, meet the terms of their responsibility contracts, and reduce per capita tax levies.[60] Other cadres devised a middle course that linked nominal enforcement to personal and collective self-interest. Rather than try to prevent excess births, they used them as a fundraising tool, collecting penalties for infractions and investing them in local projects. In one report, a village party secretary was quoted as saying that in "four or five days' time," the village had collected 10,000 yuan, "just enough to pave a village road." The reporter condemned this practice of encouraging "fundraising births" as an "evil embryo" (*guaitai*) that should be terminated, but the problem was not simply one of cadre deviance at the village level. In some areas, the money-starved family planning bureaucracy was dependent on the collection of fines to pay salaries and bonuses.[61] This preoccupation with fundraising went even further: mobilization campaigns intended to prevent excess births were perverted into fundraising campaigns, with local cadres assigned targets for the collection of fines, and with campaigns timed to occur when peasants were flush with cash.[62] Paradoxically, campaigns thus became the vehicle by which routine bureaucratic work could be sustained, giving the bureaucrats themselves a direct interest in perpetuating mobilizational methods.

In short, during the 1980s the requirements for effective mobilization were undermined by the successful implementation of other elite goals. Deepening problems with routine implementation and enforcement, however, elicited new demands for mobilization from unusual sources. In family planning, these demands did not come from conservatives threatened by the reforms, but from reform-minded experts and observers who feared the economic impact of population growth.

In 1988 and 1989 their anxieties broke to the surface. The revelation, in late 1988, that the 1987 birth rate was significantly higher than originally reported by the State Statistical Bureau, that third or additional births continued to account for as much as 17 percent of annual births, and that urban couples were also pushing for additional births, caused growing apprehension over inadequate birth control efforts and the potential "tragedy" of failing to control population growth.[63] In Hebei,

[60] Qu Yibin, "Woguo renkou chushenglu mingxian huishengde yuanyin ji duice tantao," [Reasons for our country's obvious increase in birth rate and inquiry into countermeasures], *Renkou Yanjiu* 2 (March 1988), pp. 54–55; Zhong Cheng, "Delegates and Members Show Concern for Family Planning," *Zhongguo Xinwen She*, April 11, 1988, in *FBIS*, April 13, 1988, p. 33.

[61] Jiang Su and Yang Shounian, " 'Jizi shengyu' buzu qu" [Do not permit the pursuit of 'fundraising births'], *Nongcun Gongzuo Tongxun* 4 (1988), 25.

[62] Interview with former township-level family planning cadre, March 1, 1990.

[63] Henan Provincial Service, January 21, 1989, in *FBIS*, January 25, 1989, p. 53.

these concerns led to the creation of an extraordinary coordinating body, the "family planning leading small group," to increase province-wide attention to family planning.[64] Such groups, which were pervasive in the 1970s and early 1980s, had been replaced in the mid-1980s by regular state organs and local family planning commissions. What was most telling, however, was an escalating debate between what might be called the state family planning establishment, on the one hand, and opposition professionals, social scientists, and party officials, on the other. The latter group argued that relaxation of the one-child policy was a mistake, and held policy makers and "personnel changes" responsible.[65] After this charge appeared in *Guangming Ribao* in March 1988, officials of the State Family Planning Commission felt obliged to respond at their national meeting shortly thereafter. *China Daily*, for example, reported:

> The policy that allows rural couples with only one daughter to have one more child after a certain period has proved reasonable and is not the reason for the rapid growth of China's population in recent years, it was agreed at a national meeting of the State Family Planning Commission. . . .[66]

Around the same time, an article in *Renmin Ribao* pointed out that the desire for a rapid push for zero or negative population growth must be carefully balanced against the problem of an aging population.[67]

The debate intensified in 1989, with *Jingji Ribao* picking up the call for a stricter birth control policy. Two articles, the first on January 10 and the second on January 24, stressed the impending population crises and asked, "How many centuries must it take before people can be awakened?"[68] The second, arguing that a "human wave is washing over China," pointed to the 1980 Marriage Law, the stimulating effect of the agricultural responsibility system on rural child-bearing preferences, and the relaxation of policy after 1984 as three key "policy faults."[69] It is the final paragraphs, however, that reveal the changed rationale for mobilization. Dismissing "gloomy and feeble propaganda and education" as useless in an environment where "people no longer fear punishment," the author blasted the family planning bureaucracy and "policy-makers" for their timidity; he concluded:

[64] *Hebei Ribao*, September 16, 1988, p. 1, in *FBIS*, October 12, 1988, p. 64.

[65] Liu Jingzhi, "Experts Concerned Are Not Optimistic about China's Population Situation, and Think that Interference by Officials Is an Important Reason Why Birth Rate Has Risen Again," *Guangming Ribao*, March 6, 1988, p. 2., in *FBIS*, March 18, 1988, pp. 14–15.

[66] *China Daily*, April 19, 1988, p. 1.

[67] Wu Canping and Du Yajun, "Several Questions on the Chinese Population," April 15, 1988, in *FBIS*, April 19, 1988, pp. 48–49.

[68] The quote is from Xie Zhenjiang, "There Is No Route of Retreat," *Jingji Ribao*, January 24, 1989, in *FBIS*, February 15, 1989, p. 35; see also Guo Xiao (fn. 58), 50–52.

[69] Xie (fn. 68), 36–37.

> As happy peace must be built on the foundation of powerful military strength, so are the people's voluntary and self-conscious actions to be formed on the basis of the legal and compulsory standard. We should justly and forcefully say that we must punish those who have turned a deaf ear to dissuasion from having additional children and from unscrupulously having additional children and that suitable coercion and control should be implemented in China's family planning.
>
> . . . Only if the country adopts effective and compulsory policies and at the same time, carries out propaganda and education, can the consciousness of the society be aroused. If we do not understand this problem and change the current situation soon, when we wake up in a number of years, we shall find that it will be too late.[70]

This message is a plea for effective mobilization. Despite its open advocacy of the use of coercion and compulsion, however, it is not a plea to re-create the disruptive mass campaign formula of the Maoist past. On the contrary, it is notable for its emphasis on enforcement of the "legal and compulsory standard," the adoption of "effective and compulsory policies," and for its denigration of "gloomy and feeble propaganda and education"—language out of character with the mobilization politics of the Maoist period. Instead of a disruptive mass campaign, the author calls for a targeted effort to enforce regulations already in place, and to create new, more effective ones where necessary. In short, it is a call for party mobilization to implement social policy by all necessary means, including force, but not for mobilization on a mass scale.

Sources of Mobilization in Chinese Politics

In contrast to the hypotheses of Lowenthal and Jowitt, I have argued that China in its postrevolutionary phase has not abandoned mobilizational methods to achieve party-defined goals. That is not to say that China remains a mobilization regime, but rather that mobilizational methods have been recast in ways that make them useful to the reformist elite. The result has been a variant form of campaign politics which I call institutionalized mobilization. A final objection that might be raised to this analysis is that the case of the one-child policy is unique, and therefore cannot be fairly projected on China's postrevolutionary politics in general. Since there is no denying that China's demographic dilemma is unique, the question is whether or not it has provoked a unique political solution.

The political trajectory of the one-child policy has been anything but exceptional. On the contrary, it has accurately mirrored the foremost

[70] Ibid., 37.

*political dilemma of the Deng regime in the 1980s—engineering rapid socioeconomic modernization within prescribed Leninist boundaries while relaxing tight political controls. Although the death of Mao in 1976 paved the way for the abandonment of the Maoist path to development and his radical revolutionary vision, Deng Xiaoping's alternative vision, which began to crystallize in 1978, was also radical and ambitious. Un-*like Khrushchev, Deng was prepared to do more than merely repudiate the political excesses of his famous predecessor. He embarked on a modernization strategy that swiftly and fundamentally reversed the deeply rooted Maoist approach. That reversal demanded a deradicalized political environment. But the degree of societal adjustment required for reform, the formidable political opposition to be overcome, and the scope of Deng's reform goals called for the skillful utilization of a familiar political process—mobilization of both party and society to break out of old political and economic routines in favor of new ones.

This effort to reorder an existing socialist society began with a denunciation of political campaigns and mass movements; but it was followed by a spate of dramatic reform initiatives that demanded full mobilization of the party machine. In the countryside, for example, there were tree planting campaigns, electoral reform campaigns, drives to reduce the collective burdens on peasants, the drive to build a socialist spiritual civilization, and, of course, the drive to decollectivize universally and to introduce household contracting. None of these mobilization drives resembled the radical mass campaigns of the Maoist era, and local party cadres sometimes lagged behind enthusiastic peasants. Nevertheless, the basic process of pushing through comprehensive reforms between 1978 and 1983 followed the Leninist, or "engineering," approach to social change—mobilization of the party and mass organizations to take the lead in implementation, and skillful and massive use of the propaganda apparatus to build support for the reformist line and to isolate political opponents.

This brings into question the presumption of a static definition of mobilization as a process of calculated disruption in pursuit of utopian ends. The purpose of mobilization during the stage of socialist transformation was the disruption of bourgeois social arrangements and routines in favor of new socialist arrangements;[71] in Deng Xiaoping's postrevolutionary regime, the purpose of mobilization is to ensure the responsiveness of local cadres to higher-level initiatives, and to gain societal compliance with party efforts to *reorder* the established institutions and routines of socialism. The party remains the instrument of mobilization and di-

[71] Jowitt (fn. 7).

rected change, but an atmosphere of orderly determination replaces that of "crisis and attack."[72] Mobilization has ceased to be synonymous with permanent revolution, but it remains an essential tool for implementing far-reaching reform while enforcing political orthodoxy. Crises of political power, like the Beijing Spring of 1989, may trigger a reversion to all or most of the revolutionary campaign methods of the past, but mobilizational methods also play important, if more subdued, roles in accomplishing ordinary policy goals on a national or local scale.

The party's continued reliance on mobilizational methods derives from the lingering memory of the institutionalized role that campaigns played during the Maoist era and the defining tendency of Leninist parties to rely on directed mobilization as the basic approach to political change and control. Beyond these institutional bases, however, the longevity of mobilization in China is rooted in structural and political factors.

A chronic and enduring problem in China's vast countryside is that the elite's demands outstrip regular bureaucratic capacities to meet them. In China, all bureaucratic trails lead to township and village leaders. Unlike the county-level bureaucratic apparatus, however, township governments lack the manpower and the skills to routinize party and government work effectively on a consistent basis. At the village level, leaders cease to be full-time employees of the state, and their lack of mobility at this level makes them permanent fixtures in their native communities. Although this organizational arrangement was essential to the transformation of rural China and the consolidation of the CCP's rule, the dual face of village leaders has always posed obstacles to the implementation of central policies that run counter to village interests. Although the leadership routinely tolerated a degree of slippage and corruption at the village level, its single most effective device for containing that slippage was the mass campaign. Social transformation was therefore only one of the motivations for the mobilizational approach to policy implementation. The other was the prevention of organizational decay at the grassroots.

With this tradition of rural mobilization so deeply rooted in the collective memory and so crucial to the maintenance of a disciplined party organization (attributes that separate the Chinese experience sharply from that of the Soviet Union), the formal repudiation of political campaigns did little to alter the behavioral response of local cadres who had a limited repertoire of leadership methods. In the short run, patterns established over decades ensured that party cadres remained highly sensitive to mobilizational cues from the center—so much so that they were

[72] David E. Apter, *The Politics of Modernization* (Chicago: University of Chicago Press, 1965), 360.

still susceptible to the political sins of overzealousness and "commandism" (exercising arbitrary and dictatorial rule).[73]

Over time, however, the decentralization and diversification of power resulting from successful reform eroded both the fear of retribution and the promise of meaningful reward necessary to mobilize society effectively, and altered the relationship of the party center to the far-flung party rank and file. As local cadres were rewarded more for economic entrepreneurship than for political discipline, and as the demands of newly empowered provincial or county officials took precedence over Beijing's instructions, old forms of political deviance gave way to new ones. "Coercion and commandism" (*qiangpo mingling*) in family planning yielded to neglect and to falsification of records; more broadly, economic commandism and leftism were replaced by corruption and bourgeois liberalization as the most dangerous forms of political deviance.

Yet mobilization campaigns did not disappear. Instead, they were utilized by central, provincial, and even county officials as a routine method for getting things done.[74] Moreover, as fiscal austerity started to impede government work in the late 1980s, mobilization became a convenient bureaucratic tool for overriding fiscal constraints by placing temporary political pressure on reluctant cadres.

Mobilization of some sort may not necessarily remain a permanent feature of Chinese politics. The perpetuation of variant forms of mobilization in postrevolutionary China, however, does bring into question the dichotomous phasing of communist state evolution implied in earlier models. The campaign method, which implies a mobilizational override of routine bureaucratic procedure, is an institutional and procedural correlate of Communist Party rule in China. Rather than dying away in the postrevolutionary era in favor of routinized, Weberian bureaucratism, institutionalized forms of mobilization have found common ground with bureaucratic process.[75] Regardless of official rhetoric, neither modernization writ large nor the death of Deng Xiaoping and other revolutionary elders is therefore likely to extinguish this form of campaign. That would require a more fundamental rearrangement of the structure of power than China's Leninist rulers have been willing to permit.

[73] Stuart R. Schram, *Ideology and Policy in China since the Third Plenum, 1978–84*, Research Notes and Studies No. 6 (London: Contemporary China Institute, School of Oriental and African Studies, 1984).

[74] For example, Michel Oksenberg reports that campaigns were utilized by county officials in Zouping county, Shandong Province. See his "Preliminary Impressions of Zouping County Level Government" (Paper presented at the Association for Asian Studies Annual Conference, Chicago, April 6–9, 1990).

[75] The implications for a single pattern of political authority and a non-Weberian form of bureaucratic evolution parallel the conclusions of Walder (fn. 9), chaps. 7 and 8.

FRAMING THE NATIONAL INTEREST
The Manipulation of Foreign Policy Decisions in Group Settings

By ZEEV MAOZ*

INTRODUCTION

DO reasonably smart, politically experienced leaders sometimes make national choices that go against their own best judgment, even though they have not been forced into such decisions by higher authorities or by powerful external powers? Indeed, yes: it is not at all infrequent that those who make foreign policy are manipulated into choices that they would not have made otherwise. Political manipulation is an attempt by one or more individuals to structure a group choice situation in a manner that maximizes the chances of a favorable outcome or minimizes the chances of an unfavorable one.[1] Manipulation of decisions differs from persuasion or other direct attempts to influence choice outcomes. Whereas manipulation focuses on structuring the group decision-making situation in a manner that assures success, persuasion entails direct attempts to influence (or pressure) individual decision makers into changing their preferences.

The main argument of the study has four components. First, political manipulation differs significantly from persuasion and other influence strategies in group decision making. Second, in some cases political manipulation may induce high-quality decisions, whereas in other cases it may cause the group to deviate considerably from the canons of rationality. Third, political manipulation is not invariably successful; certain situational and structural conditions contribute to its effectiveness, and others limit it. These are explored below. Fourth, even tactics of political manipulation that are skillfully applied can be effectively countered. A

* I wish to thank Steven Brams for his perceptive comments on earlier drafts. Thanks are also due to the National Security Archive, Washington, DC, for making available the transcripts of the Executive Committee meetings during the Cuban missile crisis.

[1] This definition is adapted from William H. Riker, *The Art of Political Manipulation* (New Haven: Yale University Press, 1986), ix.

World Politics 43 (October 1990), 77–110

discussion of political manipulation must, therefore, also explore some evasion tactics and countermeasures.

The essay is organized as follows. First, a theoretical background for the study of political manipulation is provided. Next, various tactics of political manipulation are discussed and illustrated by a number of case studies of crisis decisions made by American and Israeli policy groups. The paper concludes with a discussion of the conditions facilitating manipulation and of some evasion tactics.

THE GROUP-INDUCED SHIFT PHENOMENON AND POLITICAL MANIPULATION

Despite their special attributes, foreign policy groups share some traits with the more typical problem-solving groups studied by social psychologists.[2] One of the key issues that has dominated research on group-level foreign policy choices was the well-known group-induced shift phenomenon.[3] This finding suggests that it is impossible to tell before group deliberations what the choice of the group will be, even if one knows which way the participants are leaning at the beginning of the deliberations and what kind of decision rule is being used. Group dynamics have a significant impact on the choice outcome.

The explanations of the group-induced shift phenomenon fall into three general categories: persuasive argumentation theories, value theories, and groupthink theory. The persuasive argumentation school claims that group discussion is an occasion for the exchange of information and persuasive argumentation. Individual preferences change in the course of group discussion because people learn new things about the decision problem and are exposed to new information, new options, new definitions of dangers and opportunities, and new criteria by which options

[2] Zeev Maoz, *National Choices and International Processes* (Cambridge: Cambridge University Press, 1990), 206–8, identifies five special characteristics of national decision groups: (1) permanence, (2) multi-issue responsibility, (3) hierarchical structure, (4) accountability to both internal and external constituencies, and (5) common history and common future. Representative studies of group decision making in foreign policy settings are Joseph De Rivera, *The Psychological Dimension of Foreign Policy* (Columbus, OH: Bobbs-Merrill, 1968); Irving Janis, *Groupthink*, 2d ed. (Boston: Houghton Mifflin, 1982); Andrew K. Semmel and Dean G. Minix, "Small Group Dynamics and Foreign Policy Decision Making," in Lawrence Falkowski, ed., *Psychological Models in International Politics* (Boulder, CO: Westview Press, 1979); and Zeev Maoz, "The Decision to Raid Entebbe: Decision Analysis Applied to Crisis Behavior," *Journal of Conflict Resolution* 25 (December 1981), 677–707.

[3] For reviews of the various approaches on the group-induced shift phenomenon, see Dorwin Cartwright, "Risk Taking by Individuals and Groups: An Assessment of Research Involving Choice Dilemmas," *Journal of Personality and Social Psychology* 20 (September 1971), 361–78; Dean G. Pruitt, "Choice Shifts in Group Discussion: An Introductory Review," *Journal of Personality and Social Psychology* 20 (September 1971), 339–60; idem, "Conclusions: Toward an Understanding of Choice Shifts in Group Discussion," *Journal of Personality and Social Psychology* 20 (September 1971), 495–510.

can be evaluated. These factors serve as a basis for revision of individual choices, often resulting in substantive preference change. This process is rational because group discussion renders individual choices more comprehensive and less biased than they would have been had the individuals made decisions on their own.[4]

Value theories assert that groups create norms and values that cause individuals to adapt, or shift, their preferences to accord with what they perceive to be the group's norms. This implies at least some degree of irrationality because the same person faced with the same choice problem might choose one way or another, depending on his or her perception of the dominant group-shared norm.

The groupthink model shares some of its features with value theories but is more dynamic and general. Groupthink postulates a concurrence-seeking pattern wherein the individual changes preferences in accordance with what he or she perceives to be an emerging group consensus or the leader's preferences. This switch may be voluntary or forced. Voluntary preference change is due to the individual's fears of being the outcast or to the desire to belong to the in-group. Forced preference change is caused by explicit pressure put by some group participants on those who express dissenting views. Permanent policy groups that meet on a regular basis come to develop common values, perceptions, expectations, and stakes vis-à-vis outside actors, all of which enhance groupthink tendencies. The groupthink theory also emphasizes non- or extrarational aspects of the decision-making process because it demonstrates that group interaction serves to constrain comprehensive and critical analysis and evaluation of choice problems.[5]

All the explanations of group-induced shifts assume substantive changes in individual preferences. These changes are due either to exposure to new information and arguments or to new considerations that are extraneous to the decision problem itself, such as promoting or adjusting to group-shared values or seeking to preserve group harmony.

There are, however, choice shifts that occur even though there have not been actual changes in individual preferences. First, individuals may

[4] Eugene Burnstein et al., "Risky Shift is Eminently Rational," *Journal of Personality and Social Psychology* 20 (September 1971), 426–71. See also Amikam Vinokur and Eugene Burnstein, "Effects of Partially-Shared Persuasive Arguments on Group-Induced Shifts," *Journal of Personality and Social Psychology* 29 (September 1974), 305–15; Martin F. Kaplan, "Discussion Polarization Effects on Modified Jury Decision Paradigms: Informational Influences," *Sociometry* 40 (September 1977), 262–71; and Martin F. Kaplan and Charles E. Miller, "Judgment and Group Discussion: Effects of Presentation and Memory Factors on Polarization," *Sociometry* 40 (December 1977), 337–42.

[5] The most important source on this theory is Janis (fn. 2). See also Irving Janis and Leon Mann, *Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment* (New York: Free Press, 1977).

act sophisticatedly; they may choose second-best options because they reason that sincere choice may result in a worse outcome.[6] Second, group-induced shifts may result from the making of formal or informal coalitions. Individuals are expected to join that coalition which maximizes their payoffs and choose in accordance with the agreed behavior of that coalition. Here, too, although the individual's preferences have not changed, his or her choices have changed in order to maximize outcomes.[7]

That there are numerous possible explanations of the same phenomenon makes it difficult to establish which best represents reality. The model that accounts for group-induced shifts in individual choices may well depend on the structure of the group, on the nature of individual preferences, on the procedures by which the group decides, and on the circumstances under which the decision is made (for example, the amount of time available and the threats and opportunities embedded in the choice situation). Yet foreign policy literature has paid little attention to political manipulation processes that are distinct from the kind of choice shifts discussed above, despite the fact that such phenomena may be quite common.

Tactics of Political Manipulation

The term manipulation has a negative connotation. It suggests that one becomes an instrument of a more sophisticated peer and therefore

[6] The sophisticated voting explanation was originally formulated by Robin Farquharson, *Theory of Voting* (New Haven: Yale University Press, 1969). Other models of sophisticated voting have been developed by Richard Niemi and Arthur Frank, "Sophisticated Voting under the Plurality Procedure," in Peter C. Ordeshook and Kenneth A. Shepsle, eds., *Political Equilibrium* (Boston: Kluwer-Nijhoff, 1982); idem, "Sophisticated Voting under the Plurality Procedure: A Test of a New Definition," *Theory and Decision* 19 (April 1985), 151–62; and Dan S. Felsenthal, Amnon Rapoport, and Zeev Maoz, "Tacit Cooperation in Three-Alternative Non-Cooperative Voting Games: A New Model of Sophisticated Behavior Under the Plurality Procedure," *Electoral Studies* 7 (June 1988), 143–61.

[7] The coalition literature in the context of political behavior is summarized in Peter C. Ordeshook, *Game Theory and Political Theory* (New York: Cambridge University Press, 1986). A note about bureaucratic politics theories is in order here. This theory is chiefly developed in Graham T. Allison, *Essence of Decision: Explaining the Cuban Missile Crisis* (Boston: Little Brown, 1971); Graham T. Allison and Morton Halperin, "Bureaucratic Politics: A Paradigm and Some Policy Implications," in Raymond Tanter and Richard Ullman, eds., *Theory and Policy in International Relations* (Princeton: Princeton University Press, 1972), 40–79; and Morton A. Halperin, *Bureaucratic Politics and Foreign Policy* (Washington, DC: The Brookings Institution, 1974). Bureaucratic politics theory presents national choices as outcomes of intragovernmental bargaining processes in which individuals make and unmake coalitions that serve their goals more than they serve a common conception of a "national interest." This theory is often taken as portraying the behavior of political leaders as nonrational; but cooperative game theory clearly suggests that such processes are eminently rational given that individuals differ in terms of their policy preferences. Irrational group behavior reflects cases wherein individuals pass up opportunities to maximize their payoffs in group settings because of group-induced pressure; this may include even those decisions that are

does not act out of free will and for internal reasons, but rather as a pawn in some external scheme. Although this is precisely what political manipulation is all about, it must be stressed at the outset that the implications of political manipulation are not always negative for the victims. Political manipulation may well give rise to outcomes that are far superior from the group's perspective to those that would have obtained without it. Nor do I argue that those who are manipulated always act in ignorance of the kind of scheme in which they participate. Group members may knowingly and willingly play the manipulator's game because it either serves their interests or provides them a safe escape from damaging self-motivated behavior. Finally, political manipulation may be attempted simultaneously by more than one member of the decision-making group; and in fact, most of the tactics of political manipulation can be countered or circumvented quite effectively, as will be shown below.

There exist two general approaches to understanding political manipulation: a rational choice approach and a cognitive approach. (The differences between them are not a matter of whether or not political manipulation is rational. Manipulation that is rational for one person in that it guarantees a victory may be irrational for another who is manipulated and therefore loses.) The approaches are distinguished according to the kind of literature upon which they rely. The rational approach is associated with the social choice literature. The cognitive approach stems from findings that emphasize the nonrational aspects of individual and group behavior.

Some preliminary remarks about group choices are in order before these approaches are discussed. First, policy groups may be more or less hierarchical. In some cases, only one political leader has the final responsibility for national choices, although group members serve as advisers. Even then, one or more of the advisers can manipulate the decision-making process by persuasion or manipulation of the principal decision maker.[8] The key political leader may wish to manipulate his or her advisers in order to create a consensus in a divided policy group. When decisions are reached through a democratic process that requires aggregating individual preferences into a single national choice, several individuals must be manipulated to choose contrary to their intention.

---

made in the name of some notion of a "national interest." This is the opposite of what the bureaucratic politics theory leads us to believe. See Maoz (fn. 2, 1990), chap. 5, for a more detailed discussion of this issue.

[8] Even with one key decision maker, advisory groups can still play crucial roles. See I. M. Destler, *Presidents, Bureaucrats, and Foreign Policy* (Princeton: Princeton University Press, 1972), and Alexander L. George, *Presidential Decisionmaking in Foreign Policy: The Effective Use of Information and Advice* (Boulder, CO: Westview Press, 1980), for reviews of the relationship between advisory systems and policy outcomes.

Second, preferences can be manipulated through a biased and partial presentation of data that makes one's preferred policy option look good and other policy options look bad. This common form of manipulation will not be covered here because it has already been discussed extensively in the literature. (See fn. 2.) Moreover, decision makers are aware of the possibility of biased information and hence often seek out second opinions.

Third, we must explicate what we mean by group-level rationality. The notion of individual rationality has a fairly precise meaning: a decision is rational to the extent that it maximizes the subjectively expected utility of the decision maker, that is, to the extent that the alternative selected from all the available alternatives offers the best chance of achieving the decision maker's goals. But in a collective choice setting, there are at least three conceptions of rationality. *Procedural rationality* focuses on the extent to which the decision-making process follows several procedures of comprehensive search for and analysis of multiple options. Group decision making is taken to be rational to the extent that members develop multiple definitions of the situation, explore multiple alternatives in a parallel manner, engage in an open process of argumentation and exchange of information, evaluate the various policy options in terms of the value tradeoffs they entail, and choose the option that reflects a weighted aggregate of individual preferences.[9]

Outcome rationality focuses on the extent to which group decisions yield favorable outcomes. Accordingly, a decision is judged in terms of its outcome, not on the basis of how it was made. A decision is taken to be rational to the extent that it gave rise to an outcome that maximized the goals of the deciding unit.[10] *Preferential rationality* examines the relationship between individual preferences and the group's choice. The degree of rationality of group decision is evaluated in terms of the extent to which the group decision faithfully reflects the preferences of its members. What that means and whether this goal is accomplishable is the subject of much research associated with social choice theory.[11]

[9] See Janis and Mann (fn. 5), 10–14; Janice Gross Stein and Raymond Tanter, *Rational Decision Making: Israel's Security Choices, 1967* (Columbus: Ohio State University Press, 1980); Michael Brecher, "State Behavior in International Crises," *Journal of Conflict Resolution* 23 (September 1979), 446–80; Irving Janis, *Crucial Decisions* (New York: Free Press, 1989), Maoz (fn. 2, 1990), chaps. 5–6; and Maoz (fn. 2, 1981).

[10] A key question in the literature on foreign policy decision making concerns the extent to which the procedural and outcome rationality perspectives correlate. See Ole R. Holsti and Alexander L. George, "The Effects of Stress on Foreign Policy Makers," in Cornelius P. Cotter, ed., *Political Science Annual* (Indianapolis, IN: Bobbs-Merrill, 1975), 6:255–319; Janis (fn. 9); Gregory M. Herek, Irving L. Janis, and Paul Huth, "Decision Making during International Crises: Is Quality of Process Related to Outcomes?" *Journal of Conflict Resolution* 31 (June 1987), 203–26; and Maoz (fn. 2, 1990), chap. 8.

[11] On the evolution of this literature, see Duncan Black, *The Theory of Committees and*

For a number of reasons, the present study relies on the procedural ationality perspective to assess the implications of political manipulaon. First, because political manipulation is a procedural device for inuencing group choices, its implications must be examined in primarily rocedural terms. Second, the implications of political manipulation are ighly ambivalent when assessed according to the outcome and prefer-ntial rationality perspectives. The outcome of a national decision is as uch a function of the choices of other states as it is of the choice of the olicy group of the focal state. Likewise, the extent to which political anipulation brings about a decisional outcome that misrepresents the preferences of individual members may be as much a function of those preferences as it is a function of shrewd structuring of the choice setting.[12]

Manipulation That Expands Group Choices[13]

Models of political manipulation share the assumption that the would-be manipulator knows how each of the other decision makers rank orders all of the available alternatives. In some cases, the manipulator may also know the factors that account for these preferences. A well-known strategy of political manipulation is agenda setting. This can take a variety of forms, although two of them are the most common: defining the choice rule and defining the sequence by which options are introduced.

Choosing the "Right" Choice Rule

A decision maker who can determine the procedure by which the group decides is in a position to affect the choice outcome without having to induce individual preference change. Consider the following example. A group of nine decision makers must choose one of four alternatives. The preferences of the individuals over the four alternatives are given in Table 1.

If the decision in the group is reached by majority vote, then any of

Elections (Cambridge: Cambridge University Press, 1958); William H. Riker, *Liberalism against Populism* (San Francisco: Freeman, 1982); and William H. Riker and Peter Ordeshook, *An Introduction to Positive Political Theory* (Englewood Cliffs, NJ: Prentice-Hall, 1973). The classical modern work in this tradition which provides the most fundamental result in this body of research is Kenneth Arrow, *Social Choice and Individual Values* (New York: Wiley, 1951). Arrow's "paradox of voting" suggests that there exists no democratic aggregation scheme for the conversion of rational individual preferences into a rational group preference.

[12] Nevertheless, I will make occasional references to the other two perspectives in the following discussion.

[13] Much of the following discussion is based on Riker (fn. 1) and William H. Riker, "The Heresthetic of Constitution-Making: The Presidency in 1787, with Comments on Determinism and Rational Choice," *American Political Science Review* 78 (March 1984), 1–16.

TABLE 1
RANKED PREFERENCES OF FOUR ALTERNATIVES BY NINE DECISION MAKERS

Decision-Making Subgroup	*No. of Decision Makers*	*Alternatives*[a]			
		a	*b*	*c*	*d*
I	three	1	2	3	4
II	two	4	2	1	3
III	two	3	2	4	1
IV	two	4	2	3	1

[a] Ranks are: 1 = best, 4 = worst.

the members can manipulate the decision by choosing the "right" decision rule. A member of subgroups III or IV can manipulate choice by recommending simultaneous decision among all four alternatives, with the group choice being the one with the most votes. This would lead to selection of alternative *d* (with four votes) if members vote sincerely (that is, they vote for their top-ranked alternatives). By contrast, the members of groups I or II cannot assure the selection of their top-ranked alternatives in a case of simultaneous decision (*a* would have three votes, and *c*, two votes). The best that members of those groups can do is to recommend a procedure that requires sequential pairwise choice (regardless of the specific order of contests). This would assure the selection of their second-best alternative, *b*, because this alternative (called a "Condorcet alternative") beats every other alternative in a pairwise contest.[14]

Switching from simultaneous voting on all alternatives to a sequential agenda setting reduces procedural rationality. With sequencing, the group considers all the alternatives, but only two at a time. The group would therefore not consider all pairwise confrontations among alternatives, because only winning alternatives from a previous round go to the next round. Hence, the order in which alternatives are introduced determines which ones are pitted against each other and which do not confront each other at all. This has an effect on the decisional outcome in cases that do not have a Condorcet alternative.

Consider the example shown in Table 2. If the selection procedure is simultaneous, there will be a tie among the three alternatives, in which

[14] The "Condorcet alternative" is named for 18th-century French mathematician Marquis de Condorcet, who defined the principle. From the perspective of social choice theory, if such an alternative exists (and as Arrow [fn. 11] showed, this need not necessarily be the case), its selection by a decision-making (or any other voting) body is taken to be a preferentially rational decision. It follows that the selection of any other alternative (for example, alternative *d* in Table 1) is nonrational from a preferential perspective.

TABLE 2
RANKED PREFERENCES OF THREE ALTERNATIVES BY THREE DECISION MAKERS

Decision-Making Bloc	*Alternatives*[a]		
	a	*b*	*c*
I	1	2	3
II	2	3	1
III	3	1	2

[a] Ranks are: 1 = best, 3 = worst.

case a tie-breaking rule is necessary. Following an agenda-type selection procedure, however, the sequence has a significant impact on the outcome. Suppose that alternatives *a* and *b* are first pitted against each other, with the winner matched with *c*. In this case, *a* defeats *b* (by the votes of blocs I and II), but is defeated by *c* in the second round (by the votes of blocs II and III). If alternatives *a* and *c* confront each other in the first round, the final decision is alternative *b* (because *c* beats *a* in the first round but loses to *b* in the second). Finally, if alternatives *b* and *c* are considered together in the first round, the final decision is alternative *a* (*b* beats *c* in the first round but loses to *a* in the second). The manipulator can get his or her top-ranked alternative by recommending an agenda wherein that alternative is not in the first pairwise contest.

There are three problems with these agenda manipulation schemes that may render them ineffective or even counterproductive for the manipulator. First, in many cases policy groups may have well-established choice rules that cannot be changed at will. Sequencing alternatives within an agenda-type selection procedure is more feasible when some decision makers can control the order of proposals on the floor in a manner that suits their preferences.

Second, the manipulated may outsmart the manipulator by choosing sophisticatedly. If allowed to determine the sequence of selection, voter I in Table 2 will start by pitting alternatives *b* against *c*, with the winner being pitted against *a*. If all vote in direct accordance with their preferences, this sequence would lead to the selection of alternative *a*. But if voter III foresees that the final outcome would be disastrous for him because *a* is his worst choice, it would be rational for him to vote for *c* in the first round, instead of voting for *b*. This will lead to the defeat of *b* in the first round and to the defeat of *a* in the second. Since voter III prefers *c* to *a*, he can outsmart the manipulation effort of voter I.

Third, it is difficult to establish a general principle that specifies when

these manipulation tactics work in the desired way without knowing the specifics of the choice setup (that is, the number of alternatives, the distribution of individual preferences, the social preference function,[15] and the amount of information available to group members regarding the preferences of other members). The only way a Condorcet alternative can be defeated is by voting simultaneously on all the alternatives; one cannot defeat it in a sequential agenda setting, regardless of the sequence one uses.[16] When all of the preferences of the various decision makers are common knowledge, sequencing can be a double-edged weapon if other members see the trap and outmaneuver the manipulator by sophisticated choice.[17]

SPLITTING THE MAJORITY WITH A NEW ALTERNATIVE

When there is a majority coalescing about an alternative that one views as disastrous, introducing a new alternative can lead some supporters of the previous majority alternative to switch to the new one. Together with some switchers from the previous minority alternative, the new alternative may become the group's choice. Another possibility is that the addition of a new alternative may split supporters of the majority alternative and result in the conversion of the previous minority alternative into one that controls a plurality.[18]

One may argue with some truth that this does not satisfy the definition of political manipulation because the introduction of a new alternative causes preference change. But those who switch their vote need not change the ranking of the previous alternatives; they may simply add the new alternative to their preference scale. Moreover, the manipulator can maximize the chances that this alternative will succeed by designing it in a manner that addresses the fears of supporters of both the majority

[15] The social preference function is the preference ordering of the alternatives by the group as a result of all the possible pairwise contests between them. For example, in Table 1 the social preference function is $b > [a > c > d > a]$, which means that although there exists a Condorcet alternative, the society does not have a transitive ordering of the remaining alternatives. All alternatives but b are in a cycle, as is the case with respect to all three alternatives in Table 2.

[16] See Hannu Nurmi, "Voting Procedures: A Summary Analysis," *British Journal of Political Science* 13 (June 1983), 181–208.

[17] One of the most significant results of social choice theory is that there exists no social choice rule that is manipulation-proof. See A. Gibbard, "Manipulation of Voting Schemes: A General Result," *Econometrica* 41 (December 1973), 587–601. Experimental studies of sophisticated choice in agenda settings suggest that people behave sophisticatedly if the conditions regarding knowledge structures are met, although the proportion of such sophisticated choices is not very high (about 30%). See Roberta O. Herzberg and Richard K. Wilson, "Results on Sophisticated Voting in Experimental Settings," *Journal of Politics* 50 (May 1988), 471–86, and Richard K. Wilson and Andrew Pearson, "Evidence of Sophisticated Voting in Committee Setting: Theory and Experiments," *Quality and Quantity* 21 (September 1987), 255–73.

[18] Riker (fn. 1), 1.

and the minority alternatives. The manipulator thus does not affect preferences directly but uses the information regarding the determinants of those preferences to form a new alternative that requires the other decision makers to redefine their preferences.

This sort of manipulation is rational because the menu of available alternatives before the group is expanded and thus requires enhanced analysis and appraisal. Such a tactic is especially effective when the group must choose between two alternatives; in such a case, a third alternative may become a Condorcet alternative, defeating the other two in pairwise contests, or it may provoke sophisticated choice that results in a choice shift.

The introduction of a new alternative is often a feasible tactic in foreign policy settings because there are generally no formal rules to prevent the introduction of new alternatives. Moreover, this tactic can be applied independently of the choice procedure used by the group. There are potential drawbacks to this tactic, however. It may be countered by sophisticated choice or by an identical tactic on the part of opponents of the new alternative, in which case it sometimes results in just the inverse of what the manipulator had hoped to accomplish.

Consider the following example. A group of nineteen decision makers must choose between policy options *a* and *c*. A straw poll in the group reveals that thirteen members prefer *c* to *a* and six prefer *a* to *c*. One of the members who prefers *c* to *a* nonetheless feels that there exists another option, *b*, which is preferable to both and which, if proposed, could split both the majority support for *c* and the minority support of *a*. Indeed, once the new option is introduced, the group is split as follows: six members have preference order *acb*;[19] seven members have the same preference order as that of the manipulator, *bca*; four members have preference order *cab*; and two members have preference order *cba*. If all choose their top preference, *b* wins with a plurality of seven votes, topping *a* and *c* with six votes each. However, the four voters with preference *cab*, who see this outcome as unacceptable, can outsmart the manipulator by switching their support to option *a*, their second-ranked policy, leading to its selection. This, of course, is the manipulator's worst outcome. Not only is that a bad outcome for the manipulator, it is an undesirable collective outcome as well, because *c* is the Condorcet alternative.[20]

[19] Read *a* is preferred to *c* and *c* is preferred to *b*. Transitivity implies that *a* is preferred to *b*.

[20] A detailed analysis of this example is given in Dan S. Felsenthal and Zeev Maoz, "A Comparative Analysis of Sincere and Sophisticated Voting under the Plurality and Approval Voting Procedures," *Behavioral Science* 33 (April 1988), 116–30.

Nonrational Manipulation

Nonrational manipulation consists of a variety of schemes for limiting the array of choices before the group, prohibiting individuals from expressing the full array of their preferences, misrepresenting information concerning the risks and benefits involved in certain options, or deliberately underestimating the feasibility of certain measures and overestimating the feasibility of others. Because these tactics are familiar to students of foreign policy, I will not elaborate on them here.[21] Those tactics of nonrational manipulation that I discuss have been largely unexplored in a group context. These tactics are the "framing" of decisions and the "salami tactic," or the breaking down of a "big" decision into several smaller ones.

Framing Group Decisions

The work on prospect theory shows that an exceedingly large number of people from different backgrounds and with different levels of knowledge and sophistication repeatedly violate the most basic axioms of rational choice.[22] One of the more conspicuous violations is the so-called "framing" effect. Contrary to the expectations of rational choice theory, it has been shown that people are generally risk-averse with respect to gains and risk-acceptant with respect to losses, that they are more risk-acceptant for losses than they are risk-averse for gains, and that the notion of what gains and losses mean depends on a given reference point that serves as a sort of baseline.

While this result is based on experiments designed to test how people make individual choices, it has important implications for the manipulation of group decisions. An individual's choice can be reversed by framing a given choice problem differently. If it is presented as a choice between gains, one will typically choose the less risky option. However, if it is presented as a choice between losses, then one will opt for the riskier option.

Since political choices often involve trade-offs, the representation of a choice problem as one that is strictly between gains or strictly between

[21] George (fn. 8) and Janis (fn. 9) provide probably the best surveys of these problems in foreign policy settings.

[22] Daniel Kahneman and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47 (June 1979), 263–91; Amos Tversky and Daniel Kahneman, "The Framing of Decisions and the Psychology of Choice," *Science* 211 (1981), 453–58; idem "Rational Choice and the Framing of Decisions," *Journal of Business* 59, special supplement (December 1986), S225–50; and George A. Quattrone and Amos Tversky, "Contrasting Rational and Psychological Analyses of Political Choice," *American Political Science Review* 82 (September 1988), 719–36.

losses is unlikely to be convincing. People would immediately realize that they are being manipulated, though they may be unaware of what kind of manipulation is going on. Yet it has been shown that even when trade-off information is explicitly given in the decision problem, the framing tactic still influences the kinds of trade-offs people make.[23] Trade-offs between risks and losses need not be suppressed, therefore, but can be framed in a manner that will produce the desired inclination toward risk or caution. Thus, if one favors a bold and risky alternative, the framing tactic requires a definition of the decision problem as a choice between risky prospects. And if one wants the group to play it safe, the tactic requires that the problem be defined as a choice between gains. To use the framing tactic, one must create a reference point that serves as the basis for defining what is a gain and what is a loss.[24]

Framing is clearly a nonrational manipulation tactic; it capitalizes on the known tendencies of individuals to deviate from the postulates of rational decision theory. This method can be applied irrespective of the choice rule used by the group and of the social preference status of the alternatives being manipulated.

The framing result suggests that preferences can be formed in accordance with the manipulator's wish without tampering with the data that serve as the basis for the formation of individual preferences. Virtually all of the group decision theories assume that individual decision makers enter group discussion with preestablished preferences. The framing result does not require this assumption. Framing can take place whether or not people have had a chance to decide which alternative they prefer prior to the group discussion. In fact, the manipulator's task might be easier if decision makers do not have preestablished preferences than it is if they do. What is tricky about the framing tactic is that people really do not realize that they are being manipulated, that is, the reliability of the data or the fairness of the choice procedures are not in question.

Framing, like the previous tactics of political manipulation, can be countered by counterframing. One interpretation of the problem facing

[23] Quattrone and Tversky (fn. 22), 727.

[24] A caveat is in order. The experimental evidence on the framing of choices is based on conditions that are fundamentally different from those that operate in foreign policy settings. First, subjects had been given "accurate" information in an experimental sense; no such accuracy can exist in foreign policy settings. Second, experimental subjects were either uniformly knowledgeable or uniformly ignorant about the subject of the decision problem. In real foreign policy settings, members can differ widely in terms of their knowledge about the issue area into which the decision problem falls. The strength of the framing result derives from the fact that choice switch takes place even when there are no cognitive or situational constraints on rationality. This suggests that under "normal" foreign policy conditions, and with the presence of constraints on rational choice, framing may be even more effective than in the laboratory.

the group can be countered by another interpretation. The debate over definition of the problem need not be over the data, therefore, but over the implications of the data. An example of this is the debate between Robert McNamara, on the one hand, and McGeorge Bundy and Paul Nitze, on the other hand, during the Cuban missile crisis. McNamara argued that the Soviet missiles did not fundamentally alter the strategic balance. The implication was thus that the United States should either do nothing or restrict itself to diplomatic pressure on the Soviets. Bundy and Nitze argued that the missiles in Cuba almost doubled Soviet strategic power and reduced the warning time in the case of a nuclear confrontation from twenty to two minutes. The implication was that the United States had to do its utmost to bring about a swift removal of the missiles.[25]

THE SALAMI TACTIC: MAKING BIG DECISIONS THROUGH SMALL STEPS

The salami tactic is based on the known tendencies of individuals to be myopic in their decision making. Policy groups often exhibit conservative tendencies that stem partly from the nature of group decision-making processes and partly from the commitment of group members to past policies. Such commitments often lead members to view sharp departures as a sign of political inconsistency or as an admission of failure regarding past policies.[26] As public policy analysts have pointed out, the chief problem with the advocacy of policy innovations is selling them to basically conservative policy groups.[27] With the salami tactic one innovates by breaking down the radically novel course of action into a series of gradual policy options. Although each change deviates only marginally from the previous policy, each one also sets the stage for the subsequent decision in the series. Instead of having one sharp departure from the previous policy, the group takes a series of steps that ultimately has the same effect as the "innovative" alternative the manipulator desired all along.

This tactic is a continuous effort at structuring a series of decisions. Each "small" decision may accord with the group's preferences and may result from careful analysis of data and considerations of costs and benefits. But if the group had known that each decision would lead to an-

[25] Roger Hilsman, *To Move a Nation* (New York: Delta Books, 1964), 195; Allison (fn. 7), 195–97.

[26] Eugene Burnstein and Michael Berbaum, "Stages in Group Decision Making: The Decomposition of Historical Narratives," *Political Psychology* 4 (September 1983), 531–61; Janis and Mann (fn. 5), 33–35; Janis (fn. 2), 271–72; Maoz (fn. 2, 1990), chap. 5; David Dery, *Problem Definition in Policy Analysis* (Lawrence: University Press of Kansas, 1984).

[27] Dery (fn. 26).

other logical extension of the policy and that these decisions, taken together, were part of a pattern whose end was undesirable, most of its members would not have supported even one decision in the chain. Thus, the failure in rationality is over the entire series. At each stage of this process, the decision makers believed that the current problem was the consequence of the immediate situation, which was induced partly by their previous choice but mostly by external factors (such as environmental conditions, other actors' choices, and so forth). They failed to see not only that they were manipulated by one of their own but also that their choices were part of a whole pattern that led precisely to the kind of outcome that they had wished to avoid.[28]

The difficulty of detecting and counteracting the salami tactic is considerable because it capitalizes on the cognitive and situational constraints on rational choice in foreign policy settings. First, most such choices are made under conditions of uncertainty. It is difficult enough to forecast the consequences of one stage following the decision, let alone to foresee the entire chain of events being set up by the manipulator. Second, the interdependent nature of national decisions allows the manipulator to attribute the pattern to the opponent's behavior. Third, the manipulator is typically at an advantage because his or her policy advocacies at each stage of the process actually exploit the group's commitment to its past policies. Because the advocated options set up a logically consistent pattern, they appeal to those who want to preserve stable policy postures. Finally, for this tactic to be effective, the manipulator must insist at each stage of the process that a given action alternative (that which the manipulator advocates) is to be contrasted with an inaction alternative; that is, once the group has been set into a pattern of action by its previous decisions, it is the inaction alternative that would be inconsistent with previous policies.

Because the hidden agenda of the manipulator is difficult to detect, countermanipulation is unlikely. Even if some opponents are aware that

[28] Studies that explore the inherent contradiction between myopic and nonmyopic decisions include Jon Elster, *Ulysses and the Sirens: Studies in Rationality and Irrationality* (Cambridge: Cambridge University Press, 1979); Thomas C. Schelling, *Choice and Consequences* (Cambridge: Harvard University Press, 1984); and Steven J. Brams and Donald Wittman, "Non-myopic Equilibria in 2 x 2 Games," *Conflict Management and Peace Science* 6 (September 1981), 39–62. A useful analogy of the implications of myopic rationality is the well-known dollar auction game. Experimental evidence of the consequences of myopic choice in this game is given in A. Teger, *Too Much Invested to Quit* (New York: Pergamon Press, 1980), and Joel Brockner and Jeffrey Z. Rubin, *Entrapment in Escalating Conflicts: A Social Psychological Analysis* (New York: Springer Verlag, 1985). Applications of this game to international politics include Barry O'Neill, "International Escalation and the Dollar Auction," *Journal of Conflict Resolution* 30 (March 1986), 33–50; and Zeev Maoz, *Paradoxes of War: On the Art of National Self-Entrapment* (Boston: Unwin Hyman, 1989), chaps. 4, 10.

the group has been subjected to a salami tactic, they may find it difi
to counter. To do that, they must force other decision makers to (
come those cognitive and situational constraints upon which the sa
tactic builds; that is, they must persuade their colleagues that the
choice would indeed set the stage for a whole sequence of moves
would lead to undesirable outcomes. In order to accomplish that,
must convey a picture that is extremely complicated and fraught
uncertainty. The manipulator, by contrast, has only to convince deci
makers to analyze the current situation, a task that requires less in
mation to digest and analyze and that involves infinitely more confid
in prediction than does long-term forecasting of a whole series of m
and countermoves.

Conditions of Political Manipulation and Some Evasion Tactic

There is a natural tendency in a group decision setting to resort
manipulation rather than to persuasion because the former is easier
stage. What we know about attitude change in policy settings sugg
that persuasion is often very ineffective.[29] Once people are cast in a si
ation wherein a certain choice becomes a natural solution to their pro
lem, however, it is more feasible to get them to make that decision.

Nevertheless, manipulation is not always feasible; nor is it always
fective. It is therefore instructive to review briefly some of the conditio
that may make political manipulation feasible. These conditions can
divided into four categories: procedural, situational, informational, a
structural.

*Procedural conditions include those formal rules for making dec
sions. Rules that establish the procedures by which the group decide
limit considerably the feasibility of agenda manipulation, because an
manipulation must adhere to these rules. For example, rules pertaining
to procedures and responsibilities for interpreting the situation may de-
fine who is in a position to manipulate choices, that is, to frame choices,
by defining the situation.*[30] *Likewise, within an agenda setting, those who*

[29] On problems of attitude change in foreign policy settings, see Robert Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976), 288-315.

[30] For example, the centralized structure of government that characterized the Nixon-Kissinger foreign policy administration provided a convenient setting for framing by the national security adviser, who was the conduit for virtually all the information on policy problems channeled to the president. By contrast, decentralized systems of consultation and management allow more opportunities for competing definitions of the situation and hence fewer opportunities for framing. See George (fn. 8).

npowered to decide the order by which alternatives are discussed
ed upon can determine the group outcomes under certain circum-
s.

uational conditions that may facilitate manipulation are typically
associated with crisis conditions, that is, an extreme threat to basic
s or an extraordinary opportunity, coupled with acute time pres-
Under these conditions, both types of nonrational manipulation tac-
an be effective. Framing is feasible to the extent that the group does
have the time to deal with multiple interpretations of the decision
lem. The salami tactic is feasible because the problems of the mo-
t preoccupy the decision-making process and prohibit analysis of the
d, long-term picture. Crisis-induced stress allows the manipulator to
oit the ad hoc tendencies of groups under pressure. Time constraints
also prohibit the application of normal procedures and rules of dis-
ion, thereby facilitating agenda manipulation as well. It may be dif-
t to use the majority-splitting tactic of manipulation, however, be-
e new options proposed in the course of a crisis require exploration
ew information and analysis of additional implications, a luxury the
up cannot always afford.

nformational conditions are clearly the most important determinants
the feasibility and success of political manipulation. Possession of
mmetrical information and/or expertise may enable some members of
group to apply various tactics of political manipulation successfully.
o types of informational asymmetries are significant here. The first
mmetry concerns information about preferences. In the extreme case,
en only one member of the group knows the preferences of all the
er members, the former is in a strong position to manipulate group
isions. This, however, is not likely. More typically, some in the group
ow more than do some of the others about the preferences of the
embers. The higher the asymmetry in preference information, the
ore vulnerable are the less informed to manipulation and the less able
resist such efforts.

The second asymmetry concerns information about the situation and
he problem. An individual who is the only source of information about
he events requiring decision and the data necessary for the evaluation
f the various options is in a unique position to frame the issues in a
nanner that suits his or her goals without actually distorting the infor-
nation or presenting it in a selective manner. Others may find it difficult
challenge the definition of the situation, especially if the data given to
e group sound credible. Under asymmetry regarding decision-relevant

the group has been subjected to a salami tactic, they may find it difficult to counter. To do that, they must force other decision makers to overcome those cognitive and situational constraints upon which the salami tactic builds; that is, they must persuade their colleagues that the first choice would indeed set the stage for a whole sequence of moves that would lead to undesirable outcomes. In order to accomplish that, they must convey a picture that is extremely complicated and fraught with uncertainty. The manipulator, by contrast, has only to convince decision makers to analyze the current situation, a task that requires less information to digest and analyze and that involves infinitely more confidence in prediction than does long-term forecasting of a whole series of moves and countermoves.

CONDITIONS OF POLITICAL MANIPULATION AND SOME EVASION TACTICS

There is a natural tendency in a group decision setting to resort to manipulation rather than to persuasion because the former is easier to stage. What we know about attitude change in policy settings suggests that persuasion is often very ineffective.[29] Once people are cast in a situation wherein a certain choice becomes a natural solution to their problem, however, it is more feasible to get them to make that decision.

Nevertheless, manipulation is not always feasible; nor is it always effective. It is therefore instructive to review briefly some of the conditions that may make political manipulation feasible. These conditions can be divided into four categories: procedural, situational, informational, and structural.

Procedural conditions include those formal rules for making decisions. Rules that establish the procedures by which the group decides limit considerably the feasibility of agenda manipulation, because any manipulation must adhere to these rules. For example, rules pertaining to procedures and responsibilities for interpreting the situation may define who is in a position to manipulate choices, that is, to frame choices, by defining the situation.[30] Likewise, within an agenda setting, those who

[29] On problems of attitude change in foreign policy settings, see Robert Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976), 288–315.

[30] For example, the centralized structure of government that characterized the Nixon-Kissinger foreign policy administration provided a convenient setting for framing by the national security adviser, who was the conduit for virtually all the information on policy problems channeled to the president. By contrast, decentralized systems of consultation and management allow more opportunities for competing definitions of the situation and hence fewer opportunities for framing. See George (fn. 8).

are empowered to decide the order by which alternatives are discussed or voted upon can determine the group outcomes under certain circumstances.

Situational conditions that may facilitate manipulation are typically those associated with crisis conditions, that is, an extreme threat to basic values or an extraordinary opportunity, coupled with acute time pressure. Under these conditions, both types of nonrational manipulation tactics can be effective. Framing is feasible to the extent that the group does not have the time to deal with multiple interpretations of the decision problem. The salami tactic is feasible because the problems of the moment preoccupy the decision-making process and prohibit analysis of the broad, long-term picture. Crisis-induced stress allows the manipulator to exploit the ad hoc tendencies of groups under pressure. Time constraints may also prohibit the application of normal procedures and rules of discussion, thereby facilitating agenda manipulation as well. It may be difficult to use the majority-splitting tactic of manipulation, however, because new options proposed in the course of a crisis require exploration of new information and analysis of additional implications, a luxury the group cannot always afford.

Informational conditions are clearly the most important determinants of the feasibility and success of political manipulation. Possession of asymmetrical information and/or expertise may enable some members of the group to apply various tactics of political manipulation successfully. Two types of informational asymmetries are significant here. The first asymmetry concerns information about preferences. In the extreme case, when only one member of the group knows the preferences of all the other members, the former is in a strong position to manipulate group decisions. This, however, is not likely. More typically, some in the group know more than do some of the others about the preferences of the members. The higher the asymmetry in preference information, the more vulnerable are the less informed to manipulation and the less able to resist such efforts.

The second asymmetry concerns information about the situation and the problem. An individual who is the only source of information about the events requiring decision and the data necessary for the evaluation of the various options is in a unique position to frame the issues in a manner that suits his or her goals without actually distorting the information or presenting it in a selective manner. Others may find it difficult to challenge the definition of the situation, especially if the data given to the group sound credible. Under asymmetry regarding decision-relevant

information, agenda manipulation may be facilitated if the person defining the situation is also able to control the sequence of options confronting the group.

Finally, structural conditions affecting political manipulation refer to the formal and informal hierarchy in the group. Here the proposition is simple: the more strict and steep the group hierarchy, the more those at the top of the pyramid can manipulate those at the bottom and the less the latter can resist the manipulation efforts. This refers to all forms of political manipulation discussed here. Strict hierarchies provide those at the top ample latitude in fixing decision procedures and agendas, opportunities to acquire information about preferences and about decision-relevant data, and considerable freedom in framing the issues. In less hierarchical groups manipulation can be more effectively countered.

Establishing Political Manipulation in History: Methodological Comments

Before the presentation of political manipulation in historical cases, some methodological problems must be discussed. First, it is very difficult to establish whether the preferences expressed by individuals in the course of group discussion are genuine or whether they were altered due to strategic considerations, to concurrence-seeking adjustments, or to persuasive augmentation. Group discussion provides the first and only opportunity for foreign policy decision makers to reveal their preferences. Second, in many cases political manipulation is indistinguishable from other types of group-induced shifts. Very often, political manipulation occurs alongside other kinds of group dynamics, such as groupthink. For these reasons, it is imperative to identify indicators of the various manipulation tactics. These indicators are based mostly on the modus operandi of groups and group members.

Empirical Indicators of Political Manipulation

The key to identifying tactics of political manipulation in historical material is to determine who suggested what at what point of the process.[31] Agenda manipulation is probably the easiest to establish. A researcher can discern clear evidence of agenda manipulation if he or she

[31] The questions of "for what purpose and with what effect" are also important, but they are extremely difficult—if not impossible—to answer with any degree of credibility. The answer to the first question requires information about the intentions of each of the group members, and the answer to the second question requires one to resort to the outcome-rationality notion.

has information about individual preferences and finds a suggestion for change in the agenda or in the decision rule made by members who, given their preferences, would have been left in a minority. More generally, debates over the procedure by which the group is to decide constitute evidence that different members are trying to affect the outcome through the procedure by which the group decides.

Majority-splitting alternatives may also be proposed by members who sense that they are left in a minority. Often this will take place after a period of debate in which the balance of forces in the group has been revealed. The timing of the suggestion of a new alternative is significant. Such a suggestion is most often presented when the group appears to be either deadlocked or leaning toward one of the original alternatives that some support with reservations as the least of all evils. Information suggesting that the newly proposed alternative is not the top preference of the person who proposed it is evidence of the presence of the majority-splitting manipulation tactic.

Framing is probably the most difficult manipulation tactic to prove in historical research because very often only one definition of the situation is presented to the group. It is thus impossible to know whether a different explication of the choice problem would have resulted in a different decision. However, there can be indirect evidence of framing, in the form of data indicating that a decision maker altered his or her definition of the situation very drastically in a case where the previous definition of the situation would still have provided a plausible interpretation. Because political leaders value consistency, dramatic alterations of the definition of the problem associated with drastic preference change suggests framing. The trick is to distinguish between genuine perceptual change that causes preference change and instrumental change in interpretation of data. One way to make such a distinction is to examine the admissibility of the previous definition of the situation in the light of current data. If current data render the previous definition of the situation implausible, then preference change is a result of sincere change in the perceptions of group members. If, however, the group is swayed by a new definition of the situation even though the old one is still plausible, framing cannot be ruled out.

Evidence for the salami tactic includes information about the existence of a grand scheme to which the group had objected. Further evidence consists of the breakdown of such a scheme into separate stages and of discussion of each stage by the group, without reference to previous stages and potential future stages.

Another caveat concerns the assessment of the implications of political

manipulation. It is impossible to tell from analysis of historical cases what the group would have decided had it not been manipulated, that is, whether the same outcome would have obtained even without manipulation. *In light of these methodological difficulties, research on foreign policy choices in group settings should focus on ruling out the possibility of political manipulation as a plausible interpretation of group decision by determining the **absence** of these traces in the historical case. The presence of these traces can do no more than suggest that political manipulation **may** have occurred, not that it is an exclusive or even the best explanation of the observed process and the resulting choice.*

The cases discussed are treated as illustrations of how the various manipulation tactics discussed herein have been used in foreign policy choices and of the new insights that this approach provides.

Manipulation That Expanded Group Choice: The Cuban Missile Crisis and the War of Attrition

Cuba, 1962[12]

The Executive Committee (EXCOM) appointed by President Kennedy to recommend a course of action to deal with the Soviet nuclear missiles in Cuba explored a total of six alternatives. The blockade alternative was mentioned briefly on the first day of meetings but was not raised as a serious alternative until after the initial discussion, when only two other policy options were left viable: the surgical air strike and a diplomatic response of some sort. The main advantage of the diplomatic option was that it entailed the smallest chance of uncontrollable escalation. But it had fundamental drawbacks: (1) the likelihood of inducing the Soviets to remove the missiles was very low, and (2) it would have portrayed the United States as a paper tiger, causing irreparable damage to the credibility of U.S. commitments in Europe and Latin America. These advantages and drawbacks were the reverse of those related to the surgical air strike alternative. The air strike would have been an unequivocal demonstration of resolve, and it would have eliminated most, if not all, of the Soviet missiles. There would have been quite a few Soviet casualties, however, and thus the likelihood of violent Soviet response was high.

[12] This discussion is based on transcripts of the audiotaped meetings of the Executive Committee during the crisis. Hereafter they will be cited as *Transcripts*. Other sources include Allison (fn. 7); Janis (fn. 2), 132–58; Elie Abel, *The Missile Crisis* (Philadelphia: Lippincott, 1966); Arthur M. Schlesinger, Jr., *A Thousand Days: John F. Kennedy in the White House* (Cambridge, MA: Houghton Mifflin, 1965); Theodore Sorensen, *Kennedy* (New York: Harper & Row, 1965); Maxwell D. Taylor, *Swords and Plowshares* (New York: W. W. Norton, 1972); and David Detzer, *The Brink: Cuban Missile Crisis, 1962* (New York: Thomas Y. Crowell, 1979).

The choice problem, it seemed, was to determine which of the two s to select. Even the ardent supporters of the surgical air strike recnized its escalatory potential, and the supporters of the diplomatic acn were aware of its deterrence-related deficiencies. Nevertheless, a nsensus of sorts was beginning to form on the air strike alternative. Those who remained opposed to that alternative needed a strategy:

> [The] anti-airstrike coalition, however, could still offer no concrete alternative, and there lay its greatest weakness. Those men favoring forceful action were so powerful and determined, it behooved those who opposed an airstrike to come up with some *acceptable* alternative (not Stevenson's "negotiations").[33]

Against this background, the introduction of the blockade alternative by McNamara can be interpreted as a case of splitting the majority with a new alternative. The fundamental drawback of the blockade was that it did not provide a direct solution to the problem of the Soviet missiles, although it did directly address the drawbacks of both the air strike option and the diplomatic one. The air strike advocates were not easily convinced; but all those who had initially favored a diplomatic response switched to the new alternative because it offered all the advantages of a move that allowed diplomatic activity as well as the advantages of a clear demonstration of resolve. The fear of a nuclear war led some supporters of the air strike to switch to the new alternative because it was not an irrevocable act.

The arguments for the blockade were of just the sort that a majority-splitting alternative exhibits: the weaknesses of other alternatives. The seeming unfeasibility of a surgical air strike must have prompted supporters of a firm reaction to strategic support of the blockade, which they considered to be superior to the only other viable option—an indecisive diplomatic move.

Let us look at the circumstances surrounding the invocation of the blockage alternative in order to identify the indicators of political manipulation. It is significant that prior to the recommendation of the blockade, McNamara had argued for a "do nothing" alternative on the grounds that the Soviet missiles in Cuba did not dramatically alter the strategic balance of power. He must have realized that his preferred option had little support in the group; most notably, the president strongly objected to it. He also must have seen that his worst alternative had substantial support in the group. Hence, the recommendation of a blockade can be seen not only as a case of group manipulation but also as one of strategic choice on McNamara's part and also on the part of ardent

[33] Detzer (fn. 32), 115–16; emphasis in original.

supporters of more moderate action such as Stevenson. As Allison [
it:

> McNamara's vision of the nuclear specter set him firmly against the ai strike. His initial attempt to frame the issue in strategic terms struc Kennedy as particularly inappropriate, given the President's problem Once McNamara realized that the President required a strong response he and his deputy Gilpatrick seized the blockade as a fallback.[34]

All of the indicators of majority-splitting manipulation are present this case. First, the identity of the proposer is clearly indicative of t manipulative nature of the alternative. McNamara realized that he w left in a minority both in terms of his interpretation of the severity of t Soviet threat and in terms of his inclination to downplay the U.S. r sponse. Second, the timing of the proposition is also significant. Alliso mentions that by Tuesday evening it seemed that the president had mad up his mind to pursue forceful action.[35] Virtually the entire Tuesday evening session of the EXCOM, with the president in attendance, was de voted to a discussion of the technical and political implications of the air strike. It was only after the president's departure and when the agenda for the Wednesday meeting was discussed that McNamara introduced the blockade option.[36] And third, the nature of the alternative as a middle course that did not address the issue of the missiles directly but left the door open to other measures suggests that it was an attempt to sway the more moderate supporters of the air strike, such as Robert Kennedy and Ted Sorensen.[37] McNamara could also rely on the strategic support of the advocates of diplomacy and thereby break the majority favoring an air strike.

While it is difficult to establish with absolute certainty that the blockade option was a case of political manipulation rather than a sincere effort to find a deadlock-breaking middle course, the combination of all three indicators of this tactic suggest that majority-splitting manipulation cannot be ruled out.

THE ROGERS INITIATIVE OF 1970 AND THE BREAKDOWN OF THE NATIONAL UNITY COALITION IN ISRAEL[38]

The Israeli-Egyptian war of attrition, which had raged since March 1969, took a dangerous turn in the spring of 1970. The Israelis embarked

[34] Allison (fn. 7), 202. See also Schlesinger (fn. 32), 802–8.
[35] Allison (fn. 7), 202.
[36] *Transcripts*, October 16, 1962, pp. 1–52.
[37] Ibid., 202–3. This, indeed was McNamara's main argument for the blockade.
[38] Sources for this analysis include Michael Brecher, *Decisions in Israel's Foreign Policy*

mpaign of deep-penetration air raids that caused numerous ca- in Egypt. The Soviet involvement in the war became more ap-, with Soviet personnel manning antiaircraft missile batteries and Egyptian planes. This escalation raised considerable concern in ington. On June 19 a proposal drafted by the U.S. secretary of state, m Rogers, was submitted to both Egypt and Israel. This proposal d for an immediate cease-fire and the start of peace talks on a "just lasting peace" under UN auspices.[39]

Initially, the Israeli government rejected the plan, which it saw as calling for total Israeli withdrawal from the territories. The expectation was that the Egyptians would reject the proposal as well. However, the Egyptian reply to the Rogers proposal was favorable. Faced with mounting diplomatic pressure from Washington, the Israelis started intense negotiations with the U.S. designed to achieve assurances that would soften the blow of the Rogers proposal. On July 24 a dispatch from President Nixon to the Israeli premier, Golda Meir, promised continued American military and economic aid to Israel and also expressed U.S. objections to a total Israeli withdrawal to the pre–June 1967 borders. This dispatch caused a major shift in the attitude of many of the Israeli cabinet ministers to the Rogers proposal. The cabinet reconvened on July 31, following a week of intense deliberations within the various parties that comprised the national unity coalition.

Prior to that meeting, it was clear that Golda Meir was going to recommend acceptance of the Rogers proposal. The Gahal Party (a right-wing coalition of the Herut Party and the Liberal Party) led by Menachem Begin was opposed to the proposal and threatened to leave the national unity government if it was accepted. Because Gahal was a minority in the Knesset, its departure would not have destabilized the parliamentary basis of support for the government. Nevertheless, several members of the government, led by the defense minister, Moshe Dayan, felt that every effort should be made to preserve the national unity cabinet. Hence, during the government meeting Dayan proposed a third alternative: to ignore the Rogers proposal and respond, instead, to the Nixon dispatch. Accepting this dispatch with some modifications might make the Israeli position more acceptable to the hawkish faction in the government.

ondon: Oxford University Press, 1974), 454–517; Dan Margalit, *Sheder mehabayit halavan* *ispatch from the White House*] (Tel Aviv: Otpaz, 1971); Avi Shlaim and Raymond Tanter, ecision Process, Choice, and Consequences: Israel's Deep Penetration Bombing in Egypt, 70," *World Politics* 30 (July 1978), 483–516; and Ezer Weizman, *On Eagles' Wings* (New ork: Berkeley Books, 1976).

[39] The full text of the proposal is cited in Brecher (fn. 38), 490–91.

Dayan's proposal seemed to many of the ministers ambiguous and confusing. It was not clear what Israel was accepting and what it was rejecting. To other ministers, Menachem Begin in particular, this proposal was seen as a feeble attempt to mask the obvious Israeli surrender to American pressure. The ambiguity was deliberate, however. Its purpose was to split both Golda Meir's majority coalition and Menachem Begin's minority coalition, and accept the U.S. plan in the form of the Nixon dispatch while keeping the government intact. It was uncharacteristic of Dayan to display a strong attachment to any proposal in the cabinet, but this time he was adamant.[40] And his insistence led to a shift of the entire majority coalition to the option he advocated. There was also an apparent split in the minority coalition.

Dayan knew that the Liberals in the Gahal Party were willing to support a qualified acceptance of the plan. And his scheme to prevent the resignation of the Gahal ministers nearly worked. The Liberals pressured Begin to go along with Dayan's proposal and were narrowly defeated in an intraparty vote. However, in order not to appear to be opposed to a cease-fire that would end the war of attrition, Begin insisted that the government vote twice, once on Dayan's proposal and once on the cease-fire. The vote on the second issue was unanimous. The vote on Dayan's proposal was seventeen ayes and six nays.

Following the decision, Begin faced another fight in his party, this time on the withdrawal of the Gahal ministers from the government. His main rival was Ezer Weizman, who was apparently swayed by Dayan's option to believe that the approach that Gahal represented had some support in the cabinet. Again, Begin had to muster all of his oratory powers to have his way.[41]

While the manipulation effort was successful in that it led to the adoption of the manipulator's option, it was a Pyrrhic victory for Dayan. Virtually everybody was aware that Dayan's proposal constituted an effort to manipulate the Gahal ministers into staying in the government. The

[40] Ezer Weizman, then the minister of transportation, recalled that "the Alignment members were very concerned about the possibility that, if Dayan would be left in the minority he wouldn't hesitate to leave the Government. A situation in which both *Gahal* and Dayan would be out of the Government scared them. Therefore they did almost everything to appease Dayan." An anonymous civil servant observed: "Dayan announced that he would only agree to insert into the body of the text of the American initiative Israel's reservations. During the Cabinet session he was very excited and didn't abstain from shouting. It seemed that this issue was very important to him. Finally, his view prevailed." Quoted by Brecher (fn. 38), 498.

[41] On the party politics aspects of this episode, see Eitan Haber, *Begin: The Legend and the Man* (New York: Delacorte Press, 1978), 280–85; Sasson Sofer, *Begin: An Anatomy of Leadership* (New York: Basil Blackwell, 1988), 90-92; Margalit (fn. 38); and Weizman (fn. 38), 269–75.

Alignment ministers were willing to go along with this, and it seemed that the Liberal ministers were inclined to go along as well. Yet Begin's persistence and uncompromising stance prevailed over the manipulation effort, forcing all the Gahal ministers to resign.

NONRATIONAL MANIPULATION: KOREA, 1950, AND LEBANON, 1982

FRAMING EFFECTS: THE DECISION TO INTERVENE IN KOREA[42]

North Korea's invasion of South Korea on June 24, 1950, caught the Truman administration unprepared in the military or intelligence sense and also, most principally, in terms of policy. Throughout the winter and spring of 1950, and amid reports of an imminent Soviet military move, members of the administration were busy making statements excluding South Korea and Taiwan from the American defense perimeter in the Far East.[43] According to these statements, the United States had no intention of committing military forces to the defense of South Korea. It was willing to help build a strong Korean army and to supply it with the weapons it needed, but there was an interagency consensus that the South Koreans would have to defend themselves if attacked; they could not rely on American forces to do the job for them.

This policy, which had developed over years following many interdepartmental consultations and budgetary allocations, changed from one extreme to the other almost overnight. The decision to intervene was nearly unanimous within the administration. One of the participants pointed out that the decision to intervene in the war revealed "the finest spirit of harmony I have ever known."[44]

The decision to intervene in the Korean War was made in two stages. The first was the June 25 decision to evacuate American citizens from Korea and to repel any attack on the airports or the roads leading to them. The second decision, to commit U.S. air and naval units to the fighting, was made on June 26. Both decisions were reached during what

[42] Sources for this analysis include Alexander L. George and Richard Smoke, *Deterrence in American Foreign Policy: Theory and Policy* (New York: Columbia University Press, 1974), 140–83; Harry S. Truman, *Memoirs: Years of Trial and Hope* (Garden City, NY: Doubleday, 1965); Dean Acheson, *Present at the Creation: My Years in the State Department* (New York: W. W. Norton, 1965); Glenn D. Paige, *The Korean Decision* (New York: Free Press, 1968); idem, *1950: Truman's Decision, the United States Enters the Korean War* (New York: Chelsea House, 1970); idem, "On Values and Science: The Korean Decision Reconsidered," *American Political Science Review* 71 (December 1977), 1,063–69; and De Rivera (fn. 2).

[43] The statements by Dean Acheson, Senator Tom Connally, and John Foster Dulles are given in Paige (fn. 42, 1970), 7–31, 42–45.

[44] *New York Times*, June 28, 1950, p. 4. Quoted in Paige (fn. 42, 1968), 179. Janis (fn. 2), 49–50, takes this as a symptom of groupthink.

came to be known as "the Blair House conferences," at two separate sessions of the president and his political and military aides. The structure of both conferences was nearly identical. The president opened the discussion by saying that he wanted to hear his advisers' views and recommendations with "an open mind"[45] and then called upon the secretary of state, Dean Acheson, to review the situation and make recommendations. Acheson gave "a darkening report of great confusion"[46] and read a list of operational recommendations, after which the president "asked each person in turn to state his disagreement and any views he might have in addition."[47] The discussions that followed were limited to two issues: the technical and practical implications of Acheson's recommendations and the likely Soviet response.

The group exhibited no dissent, and no other options were suggested. Nor was a "do nothing" option discussed explicitly, though it loomed large as an implicit alternative to Acheson's recommendations.[48] It is difficult to prove that either the president or the secretary of state was engaged in a deliberate plan of political manipulation, but this possibility cannot be rejected out of hand. Several factors suggest that the participants in the discussion may have been victims not only of groupthink but also of a framing tactic of political manipulation.

First, the composition of the advisory group is important. The fifteen men assembled at the conferences in addition to the president were all in charge of foreign policy and national security affairs. None of the individuals present were in charge of domestic political or economic affairs. As a result, the discussion of economic and political implications of the decisions was clearly lacking. The impression of "open mind" that Truman wanted to convey to the group was not evident when Undersecretary of State Webb wished to discuss the domestic political aspects of the situation. The president snapped: "We are not going to talk about politics. I'll handle the political affairs."[49]

Apparently, both Blair House conferences were guided by the dominant definition of the situation and the policy recommendations of Dean Acheson. By conferring the role of the official interpreter of the situation and the responsibility for policy recommendations on one person, the

[45] Paige (fn. 42, 1968), 126.

[46] Acheson (fn. 42), 406.

[47] Truman (fn. 42), 334.

[48] Truman points out that there "was no suggestion from anyone that either the United Nations or the United States could back away from it. This was the test of all the talk of the last five years of collective security." Ibid. Secretary of Defense Johnson stated that during the second Blair House conference, "if we wanted to oppose [direct military intervention], then was our time to oppose it. Not a single one of us did. There were some pointing out of the difficulties . . . and then the President made his decision . . . which I thought was the right decision." Quoted in Paige (fn. 42, 1968), 179.

[49] Quoted in Paige (fn. 42, 1968), 141.

president in effect set an agenda wherein Acheson's proposals were seen as having the president's stamp of approval. Moreover, in both conferences the defense department did not have policy recommendations of its own, though Secretary Johnson later recalled that "neither I nor any member of the Military Establishment in my presence recommended we go into Korea. The recommendation came from the Secretary of State, but I want to repeat that it was not opposed by the Defense Department all the members of which had severally pointed out the trouble, the trials, tribulations, and the difficulties."[50]

Because the definition of the situation was so gloomy, it was difficult to oppose a unique action alternative. From the outset, the North Korean aggression was presented as having been planned and inspired by the Soviet Union. Seen in this light, the issue was not whether to intervene in a local civil war, as had been the issue for American policy prior to June 24, but rather whether to resist a major Soviet test of American resolve. This is clearly manifested in how the decision makers dealt with Soviet intentions. Not once was the possibility raised that the North Koreans were acting without Soviet knowledge or approval. Virtually all interpretations of Soviet aims implied that the Soviets would have been imprudent to absent themselves from the United Nations Security Council meetings, yet that was what they did. This was explained as being a consequence of delays in the Soviet decision-making process or of a failure on the Soviet part to foresee UN action.[51]

The perception that the invasion was a Soviet ploy to test American resolve switched the reference point to one defining the matter as a grave threat to the national interest. Failure to resist the Soviet-inspired aggression could result in additional Soviet challenges of a more severe and frequent nature, up to a general war. Resistance to this aggression could lead to uncontrolled escalation, but even if it did not, it would certainly result in numerous American casualties. That they may have selected an extremely risky course of action did not escape the decision makers; the dominant view was that it "was a calculated risk."[52] There was even an explicit readiness on the part of both civilian and military participants to escalate the war to a general one.[53] The combined effect of a dichotomous

[50] Secretary Johnson's testimony before the Committees on Armed Services and Foreign Relations. Quoted in Paige (fn. 42, 1968), 164.

[51] Ibid., 132; Acheson (fn. 42), 408. See also Alexander L. George, "American Policy-Making and the North Korean Aggression," *World Politics* 7 (January 1955), 209–32, for additional interpretations of Soviet intentions and their implications for U.S. policy.

[52] Secretary Johnson's testimony, quoted in Paige (fn. 42, 1968), 173.

[53] General Bradley, the chairman of the Joint Chiefs of Staff wrote several months later: "If the defense of South Korea was risking an all-out war the choice was not ours for the Communists had thrown down the gauntlet." Phillip Jessup, the UN ambassador, argued: "The invasion had to be met even if it meant the beginning of World War III." Ibid.

choice agenda, the presentation of a policy recommendation as if it ha
come straight from the president,[54] and the framing of the problem as
gamble involving risky prospects made the choices that followe
straightforward courses of action and led all participants to believe tha
they were serving the best interest of the United States.

But all this is not enough to suggest that a deliberate manipulation attempt took place. One may rightfully ask whether a different interpretation of the problem had been possible, and whether—given such an interpretation—a different course of action would have been adopted. Although speculation on that which did not happen is a risky venture, available evidence suggests that other ways to interpret the problem were more than reasonable. In fact, the administration's policy prior to the invasion had been based on an entirely different reference point—the interpretation of the conflict as fundamentally civil, or of little inherent strategic interest to the United States. The situation could easily have been presented by participants at the Blair House conference as it had been presented by Senator Tom Connally in his May 5 interview,[55] namely, that Korea was not a major strategic asset and that the United States should limit itself to furnishing the Koreans with sufficient military and economic aid to resist the communists on their own. The choice could thus have been between continuation of the previous policy toward Korea and military intervention rather than one between military intervention and doing nothing. However, this would have required setting the existing U.S. policy toward Korea as the reference point.

Here, too, the indicators of framing are clearly present. First, the people involved transformed their definition of the situation overnight. The change in the perception of the problem from that of a civil conflict to a Soviet test of U.S. resolve changed the inclination of the group to take risks. Not only was the definition of the situation changed with regard to Korea, but the decision to send the Seventh Fleet to the Formosa Straits suggests that the problem was framed as a general communist ploy involving both the Soviet Union and Red China. Second, the evidence of Soviet involvement in the North Korean invasion available to U.S. policy makers at the time was far more tenuous than the evidence consistent with the preinvasion framing of the problem. Yet no one challenged the dominant Truman-Acheson frame of a Soviet test of American resolve. Although it is obvious that all of the symptoms of the group-

[54] For example, Secretary of Defense Johnson claimed that he had first heard of the recommendation for direct military intervention from the president prior to the second Blair House conference. Ibid., 164.

[55] Quoted in Paige (fn. 42, 1970), 30–31.

think phenomenon were present, the concurrence-seeking pattern of the members of the Blair House conferences was induced by a change in the administration's framing of the problem. Moreover, a groupthink explanation cannot account for the connection between the new framing of the problem as a choice between losses and the policy recommendation of intervention.

SALAMI TACTICS IN POLITICAL DECISION MAKING: THE LEBANON INVASION OF 1982

It is often difficult to detect a master plan in political and military affairs that does not lend itself to multiple interpretations. One of the reasons for this is that decision makers simply do not operate on the basis of grand strategies and master plans; rather, they tend to act as problem solvers and fire extinguishers.[56] The planning and execution of the Israeli invasion of Lebanon is a clear exception in this sense. The political goals of the invasion, its scope, and its precise outline had been put down on paper many months prior to the actual invasion.[57] We therefore know quite well what the goals of the manipulator had been, and we have a good notion of what the invasion was supposed to look like. As envisaged by Ariel Sharon, the Israeli defense minister, the goals of the war were fourfold: (1) to destroy the PLO political and military infrastructure in Lebanon, including the physical elimination or forceful expulsion of the guerrillas from Lebanon, (2) to push the Syrian army out of Lebanon, and (3) to establish a strong Christian-controlled central government in Lebanon that would (4) sign a peace treaty with Israel.[58]

Such far-reaching objectives were unprecedented in Israel's foreign policy, and they carried major operational implications. First, the objectives necessitated occupation of Beirut, Lebanon's capital. Second, they required a major confrontation with the Syrians, with all the attendant risks of escalation. Third, the financial costs and the number of casualties would be considerable. So would be the adverse diplomatic repercussions of a war that could not be construed, even by the wildest stretch of the imagination, as defensive.

[56] Maoz (fn. 2, 1990), chap. 9.

[57] The most explicit delineation of the political goals of the invasion is given in the Sharon-Gemayel agreement of January 1982. The scope and form of the invasion had been set forth in the operational plan prepared by the Israeli army and was given the name Big Pines by he army computer. The plan was presented to the Israeli cabinet in an informal session on December 20, 1981. It was flatly rejected by a majority of the ministers. See Zeev Schiff and Ehud Yaari, *Israel's Lebanon War* (New York: Simon and Schuster, 1984), 47–48, and Avner Yaniv, *Dilemmas of Security: Politics, Strategy, and the Israeli Experience in Lebanon* (New York: Oxford University Press, 1987), 107.

[58] Schiff and Yaari (fn. 57), 45–54; Yaniv (fn. 57), 100–104.

That Sharon was aware of these implications was obvious from the nature of the Big Pines war plan. This plan called for the frontal movement into Lebanon of two armored columns from the south, as well as for two landings of Israeli forces from the sea in Sidon (about halfway between the Israeli-Lebanese border and Beirut) and Junieh (a city north of Beirut), as well as landings by airborne forces in key strategic positions along the Beirut-Damascus highway. The plan also required immediate confrontation with the Syrian forces and their isolation from the Syrian-Lebanese border before the Syrians could rush in reinforcement. In effect, the Big Pines plan envisioned strategic envelopment of both the PLO and the Syrian forces.

Of course, both the political aims and the military plan were overly ambitious, as Sharon was to learn. On December 20, 1981, in response to a request for a plan for a large-scale attack on the PLO in Lebanon, he presented the Big Pines plan. To Sharon's (and Begin's) dismay, however, this proved too much for the government. Realizing that there was no support in the government for this plan, Sharon modified its scope several times. New versions of the plan were presented to the cabinet, and each time, the scope of the proposed invasion was more modest than in previous plans.[59]

Following an attempt to assassinate the Israeli ambassador to London, the Israeli government approved a much more modest invasion plan. This plan stipulated that the invasion was to be limited to a twenty-five-mile zone north of the Israeli border and that the Israeli forces would make every effort to avoid a fight with the Syrians. But the plan the government had supported would not have accomplished any of the objectives set forth in the Sharon-Gemayel agreement of January 1982. Hence, Sharon's next move was to maneuver the army into a position wherein a clash with the Syrians would be unavoidable. A day after the war started, Sharon

> confronted the cabinet . . . with a request to outflank the Syrians in a manner that would force them to retreat but also ultimately take the IDF to the Beirut-Damascus road many miles beyond the twenty-five mile limit

[59] Yaniv (fn. 57), 107–9, reports at least five such abortive attempts to mobilize governmental support for any sort of invasion: on January 28, twice in March, and on April 11, 1982. At the last of these meetings, the government authorized a large-scale aerial strike on PLO bases. Sharon and Begin hoped that such a strike would provoke the PLO into significant shelling of Israeli villages in the north, which would cause the dissidents in the cabinet to switch choices. This, however, did not happen. Hence they had to wait patiently another two months. Another such meeting took place on May 10. At this point, while there was a majority support for the limited plan, seven out of eighteen ministers, including the two deputies of the prime minister, Simcha Ehrlich and David Levy, voted against it. The plan was again shelved and modified.

> which the cabinet had initially approved. The dilemma, of which Sharon may have been fully aware before, was presented to the cabinet when the war was already in progress. There was no going back, and therefore the cabinet had no logical alternative but to authorize the move. In a vain attempt to prevent turning the outflanking maneuver into a general confrontation with the Ba'ath republic, the cabinet demanded that an approach be made toward Syria promising to avoid confrontation if the Syrian forces would not open fire.[60]

At the point, military considerations took over. Sharon appeared reg-ılarly at cabinet meetings requesting authorization to confront Syrian orces that were threatening the advance of the Israeli forces. Each such uthorization was deemed necessary because the outflanking maneuver vas performed on a route that placed the Israeli troops in a topographi-ally inferior position to Syrian forces. Sharon's argument was that if the yrians decided to attack, they could use their superior positions to inflict ıigh casualties on the Israeli forces. The cabinet had no choice but to pprove the recommended measures.

Ultimately, a number of cabinet ministers realized that they had been he targets of a clever manipulation process wherein they authorized tep-by-step the very same plan they had opposed all along. Sharon's equests for the authorization of additional military moves met stiffer pposition. At this point, Sharon resorted to framing. As Yaniv points ut, "He was thus thrust into the position of a losing gambler: to stop vould mean sustaining an assured loss; to carry on might increase the oss but also the possibility of an ultimate success that would justify all he costs."[61] But this was not only Sharon's personal problem: virtually ll the ministers who had played along with Sharon found themselves n the horns of the same dilemma. And this was precisely how the issue vas presented to the cabinet during the siege of West Beirut. When re-uesting authorization for the massive bombing of West Beirut, Sharon ramed the problem as a choice between the loss of what had been ac-omplished thus far (and giving the PLO a major political victory) and he criticism of the entire world following the shelling of a populated ity. The cabinet again had to go along with the plan.[62]

[60] Ibid., 113.

[61] Ibid., 115.

[62] Sharon had consistently authorized the army to advance and perform operations that ad no cabinet approval and that would almost certainly have been turned down by the abinet. Yet precisely these moves forced the cabinet to approve other measures that violated he members' better judgment. For example, on June 21–22, 1982, Sharon ordered the army o improve its positions along the Beirut-Damascus highway and in the outskirts of West eirut, despite a previous cease-fire agreement. This was done without cabinet approval and esulted in numerous Israeli casualties. Once the cabinet learned of these incursions, Sharon vas promptly ordered to cease fire. He refused. A special cabinet meeting was called on June

The Lebanon case illustrates the decision makers' heightened vulnerability to manipulation in times of crisis-induced stress. Each of the decisions the Israeli cabinet made during the war had an extremely short deadline attached to it. Options that were raised by ministers other than Sharon were rejected out of hand because there was no time to analyze their implications. The most extreme cases were the decisions to enter West Beirut and to send the Phalange militia into the refugee camps of Sabra and Shatilla, the latter of which resulted in a brutal massacre.[63]

It is difficult to ascertain the extent to which Sharon was the sole mastermind of the manipulation campaign. Opinions are divided on whether Prime Minister Begin was an accomplice or a victim. What is clear, however, is that this process of manipulation brought about the first defeat in war in Israel's history. The prolonged and painful Israeli experience in Lebanon was a direct consequence of a flawed decision process, during which experienced politicians were pawns in the hands of one of their colleagues.[64]

Conclusion

The cases discussed above demonstrate several things about the process and outcomes of political manipulation. First, manipulation may be used by those in the minority who face the prospect of having their worst alternative selected by the policy group. This was the case both in the Cuban missile crisis and in the Lebanon War. But it can also be used as a tactic by powerful group members who want to generate widespread support for a policy that could otherwise become a source of contention. This was demonstrated in the Rogers initiative and in the Korean decision. Second, several tactics of political manipulation may be employed

---

24, but "since by then the operation was nearly completed and much blood had already been spilt, Sharon once again succeeded in prevailing upon his weak and bewildered colleagues. The cabinet at last approved continuation of the attack on Alei-Behamdoun, and it was completed two days later." Ibid.; Schiff and Yaari (fn. 57), 203–7.

[63] On the implications of these manipulations for the outcome of the Lebanon War, see Zeev Maoz, "Power, Capabilities, and Paradoxical Conflict Outcomes," *World Politics* 41 (January 1989), 239–66, and Maoz (fn. 28), chap. 10. It must be noted that manipulation of the government was suspected all along and opposed by key people both in the army, for example, the chief of military intelligence, General Saguy, and in the Labor Party, whose leaders were informed of the imminent invasion and repeatedly expressed reservations regarding its proposed scope and general wisdom. See Schiff and Yaari (fn. 57), and Aryeh Naor, *Memshalah bemilhama* [Cabinet at war] (Tel Aviv: Lahav, 1986).

[64] On this decision, see Zeev Maoz, "Multiple Paths to Choice: An Approach to the Analysis of Foreign Policy Decision," in Irmtraud N. Gallhofer, Willem E. Saris, and Marianne Melman, eds., *Text Analysis Procedures for the Study of Decision Making* (Amsterdam: Sociometric Research Foundation, 1986), 69–96. In the Cuban missile crisis, though time pressure was definitely a factor, the participants in the EXCOM had, relatively speaking, more time to explore multiple courses of action.

taneously. In the Lebanon case, the manipulator employed framing and sequencing at the same time. Another interesting example is the introduction of the blockade alternative by McNamara during the Cuban missile crisis. The following paragraph illustrates very nicely the linkage between framing of the decision problem and the advocacy of a policy option.

> MCNAMARA: Now, the second alternative I'd like to discuss just a second, because we haven't discussed fully today, and I alluded ... to it a moment ago. I'll be quite frank. I don't think there is a military problem here. This is my answer to Mac's [McGeorge Bundy's] question ... and therefore ... I asked myself, Well, what is it then if it isn't a military problem? Well, it's just exactly *this* problem, that ... if Cuba should possess a capability to carry out offensive actions against the U.S., the U.S. should act. ... This is a domestic political problem. The announcement—we didn't say we'd go in and ... kill them, we said we'd *act*. Well, how will we act? Well we want to act to prevent their use, and it's really the ... act. Now, how do we ... act to prevent their use? Well, first place, we carry out open surveillance, so we know what they're doing. All times. Twenty-four hours a day from now and forever, in a sense indefinitely. What else do we do? We prevent any further offensive weapons coming in. In other words, we blockade offensive weapons.[65]

The blockade option was introduced not only as a majority-splitting alternative but also as a solution to a definition of the problem as a challenge to U.S. commitment and its willingness to carry out threats.

Third, political manipulators can take advantage of situations or manipulate situations so that their tactic is effective. The Lebanon case shows that setting short deadlines for each decision in the sequence can be an effective mechanism for obscuring the pattern of decision into which group members are being cast. Thus, not only is political manipulation more feasible under certain situations, but manipulators may actually structure the situation to their advantage.

Political manipulation may have its pluses, certainly for the manipulator, but also for the manipulated. The present study does not attempt to relate the procedural rationality of the manipulation tactic to the substantive quality of the decisional outcome. Whether such a relationship exists is left for future research to determine. Even in the absence of such a relationship, however, decision makers must be aware of the fact that a certain decisional outcome has been induced by shrewd manipulation tactics and that, without those tactics, another outcome might have obtained. When political manipulation induces choices that counter the social preference, it is important that the group know, first, that it is being manipulated, and, second, how to evade or circumvent these tactics.

[65] *Transcripts*, October 16, 1962, pp. 45–46.

Some general rules apply. One is that awareness of an attempt at political manipulation is the first step in rendering it ineffective if one so wishes. Second, sharing that knowledge with other group members can be a good way of reducing the effectiveness of the political manipulation tactic. Third, when the manipulator is powerful in some respects that are crucial to the success of his or her tactics (for example, when the manipulator has the formal authority to set the agenda), the best antidote for one political manipulation tactic is another political manipulation tactic. Fourth, the four tactics of political manipulation capitalize on the tendency of people to choose sincerely, that is, in direct accordance with their preferences. Hence, sophisticated choice via tacit or explicit coalition formation provides an effective tool to counter political manipulation.

In addition to the evasion tactics mentioned above, the suggestions of various scholars regarding other remedies of defective group choices also apply. In particular, multiple advocacy, open discussion, and multiple channels of information[66] do more than safeguard against other forms of defective decisions that are due to the structure and dynamics of group decision making. They can also alert decision makers to manipulation efforts and help them resist those efforts that can harm the aims pursued by the group.

This essay barely scratches the surface of the study of political manipulation in foreign policy decisions. The topic has been underresearched, and the purpose of this study is to highlight some of the possible avenues for additional work in this area. Future research might better explore the differences between political manipulation and other forms of personal influence in policy groups, additional tactics of manipulation, conditions facilitating such tactics, and prevention measures.

[66] See in particular George (fn. 8), 139–216; Janis (fn. 2), 260–76; Janis and Mann (fn. 5), 367–404; Daniel D. Wheeler and Irving L. Janis, *A Practical Guide for Making Decisions* (New York: Free Press, 1980); Yehuda Ben-Meir, *National Security Decisionmaking: The Israeli Case* (Boulder, CO: Westview Press, 1986), 129–44; and Janis (fn. 9).

THE STATE AND THE EFFECTIVE CONTROL OF FOREIGN CAPITAL
The Case of South Korea

By RUSSELL MARDON*

I. Introduction

THE pattern of industrial development in the Republic of Korea (hereafter referred to as Korea) offers several important insights ıto the broader theoretical debate on the political economy of develop- ıent. Among these is a significant issue regarding the integration of ›reign capital into the process of domestic industrial development and s effect on domestic entrepreneurial activity. The Korean case offers ısight on this issue because Korea is one of the few developing countries ·tually experiencing a diversified industrialization process; more im- ›rtantly, it does so because in this industrial development the means of ·oduction are predominantly owned by domestic actors, and local mar- :t shares are predominantly controlled by domestic producers. This ıttern of development is significant in that it is not predicted and cannot : adequately explained within the framework of either of the dominant hools of thought on economic development: the neoclassical and the :pendent development school.

Neoclassical analysis views foreign capital and technology as a precon- ıtion to the initiation and maintenance of an efficient industrialization rocess. Its proponents argue that developing economies which are open › foreign investment can receive the benefits of foreign capital, technol- ɡy, and market expertise that are absent from domestic sources. They ·e foreign ownership of the means of production as a positive force that ads to industrial growth and to the integration of the developing econ- ny into the world market.

Neoclassical studies focused on Korea have generally attributed in- ustrial growth to correct policy choices that effectively oriented the do- ıestic economy to international market forces.[1] To attribute Korea's

* I would like to thank Robert Jackman for his comments and suggestions. I would also ‹e to thank the Department of Political Science, Michigan State University, for assistance I ·ceived while conducting the early stages of this study.

[1] See Bela Balassa, ed., *Development Strategies in Semi-Industrial Economies* (Baltimore: The

economic success to market forces alone would be to ignore the highly active role of Korean political institutions in comprehensive sectoral planning, in controlling the pattern of investment, in providing subsidies, monopoly rights, and market protection to domestic producers, and its strict regulation of foreign capital inflows. The neoclassical model cannot fully specify an industrial development experience led by high levels of state guidance over primary economic activities.

Theorists of dependent development agree that foreign capital and technology can lead to industrial growth in a developing nation. In order to attract the level of foreign capital and technology that is requisite for industrial development, however, political structures must orient economic policy to the interests of foreign capital.[2] This, in turn, leads to an absence of domestic industrial entrepreneurship, as well as to foreign ownership of the means of production, foreign control of the domestic market, and the expropriation of national surplus.

The dependent development explanation is severely limited when applied to the case of Korea, where foreign capital has not been the primary beneficiary of industrial growth. The levels of foreign investment and foreign control of domestic markets are extremely low when compared to those of domestic investment. Firms such as Hyundai, Samsung, Daewoo, and Gold Star have dominated production and monopolized domestic market shares in nearly all sectors. Nationally controlled capitalist development of this scope and magnitude at the periphery was never envisioned by the dependent development school.

The logic of both neoclassical and dependent development predicts—although opinions differ on the net effect of benefits derived—that where industrialization does occur in the developing world, it is characterized by foreign ownership and foreign dominance of local markets. Among those developing nations that have experienced industrial growth, however, the empirical pattern of foreign ownership varies greatly, as does the pattern of foreign control over local markets. This suggests that variations of domestic political structures and the imple-

Johns Hopkins University Press, 1982), and *The Newly Industrializing Countries in the World Economy* (New York: Pergamon, 1981); Edward K. T. Chen, *Hypergrowth in Asian Economies* (London: Macmillan, 1979); Larry Westphal, "The Republic of Korea's Experience with Export-Led Industrial Development," *World Development* (June 1978), 347–82; World Bank, *World Development Report* (New York: Oxford University Press, 1983); Anne O. Krueger, *The Developmental Role of the Foreign Sector and Aid* (Cambridge: Harvard University Press, 1979).

[2] See Peter Evans, *Dependent Development: The Alliance of Multinational, State, and Local Capital in Brazil* (Princeton: Princeton University Press, 1979); Fernando Cardoso, "Associated Dependent Development," in Alfred Stepan, ed., *Authoritarian Brazil* (Princeton: Princeton University Press, 1979); Guillermo O'Donnell, *Modernization and Bureaucratic Authoritarianism: Studies in Latin American Politics*, Politics of Modernization Series No. 9 (Berkeley: Institute of International Studies, University of California, 1973).

mentation of alternative development strategies may have an important influence on establishing a country's linkages with the international economy and on foreign economic integration into the developing economy.

Recent studies have systematically examined Korean political structures and institutions in relation to economic performance. They document that government intervention extends deeply into several key areas of economic activity.[3] It is becoming increasingly clear that industrial development in Korea can be properly specified only when attention is focused on political as well as economic factors. These findings, and the inability of leading theoretical perspectives to explain the pattern of foreign capital integration into the Korean economy, pose an interesting set of questions. Are domestic political factors largely responsible for shaping the pattern of foreign capital integration into the Korean industrialization process? And if so, did this pattern of foreign capital integration facilitate the growth of domestic industrial entrepreneurship and control over the local market?

The present study attempts to answer these questions. It focuses on the governmental strategies and mechanisms utilized to shape a pattern of foreign economic inflows into Korea that provided the capital and technology necessary for industrial development but minimized the level of foreign ownership and market control. Prior to an examination of legal and bureaucratic instruments and their interaction with international economic forces, it is necessary, however, to understand the significant level of political autonomy enjoyed by Korean political leaders during the process of industrial development. This autonomy has allowed the Korean government to pursue an economic policy relatively independent of domestic social forces and international pressures.[4] Without this high level of autonomy, the nationalistic economic policies described below could not have been designed and implemented consistently.

[3] See Leroy Jones and Il Sakong, *Government, Business, and Entrepreneurship in Economic Development: The Korean Case* (Cambridge: Harvard University Press, 1980); Chalmers Johnson, "Political Institutions and Economic Performance: The Government-Business Relationship in Japan, South Korea, and Taiwan," in Frederic C. Deyo, ed., *The Political Economy of the New Asian Industrialism* (Ithaca, NY: Cornell University Press, 1987); Stephan Haggard, Byung-kook Kim, and Chung-in Moon, "The Transition to Export-Led Growth in Korea," presented at Conference on the Role of the State in Economic Development: Republic of Korea (University of California, Los Angeles, August 14–16, 1987); Young-Il Lim, *Government Policy and Private Enterprise: Korean Experience with Industrialization* (Berkeley, CA: East Asian Institute, 1981).

[4] Stephan Haggard and Chung-in Moon, "The South Korean State in the International Economy: Liberal, Dependent, or Mercantile?" in John Ruggie, ed., *The Antinomies of Interdependence* (New York: Columbia University Press, 1983), 140; Chalmers Johnson (fn. 3), 151; Hagen Koo, "The Interplay of State, Social Class, and World System in East Asian Development: The Cases of South Korea and Taiwan," in Deyo (fn. 3), 175.

Three political conditions served to insulate the state's decision-making capacity from external influence. First, when the military government took power in 1961, the powerful social groups who traditionally tend to dominate political power in Third World countries did not exist in Korea. Land reforms in 1947 and 1950 had virtually eliminated the landed class.[5] A merchant class whose power was tied to its ability to gain access to aid funds and government-issued import licenses began to emerge in the 1950s;[6] but, with the fall of Syngman Rhee in 1960, the source of its economic and political power was virtually eliminated.

A second precondition was the absence of entrenched foreign investment and other established ties to the international economy. With the defeat of Japan, all linkages between foreign capital and the Korean comprador elite had been severed. In 1961, there were no foreign investments in place in Korea, and less than 1 percent of GNP was earned through exports.[7] In countries where the ruling elites are in a comprador position—that is, where they have established personal wealth and vested interests interlinked with foreign capital—the state does not possess the necessary capacity and commitment to organize governmental mechanisms for the effective regulation of that foreign capital. The Korean leadership of 1961 had no significant ties to foreign capital. Rather, the new military leaders perceived that their personal political and economic interests, as well as those of the state, would be best served by the development of industrial production based upon government-directed sectoral planning. A domestic industrial class could be expected to follow the government's economic guidance through state control of banks and investment funds while foreign capital could not be expected to do so. The government was therefore strongly committed to nurture the development of domestic firms and restrict the inflow of foreign capital.

A third precondition that enhanced the capacity of the Korean government to implement a nationalistic policy toward foreign capital was the country's strategic relationship with the United States. Even though Korea discriminated against American capital and goods, Korean producers had virtually unlimited access to the American market as well as extensive access to American and international aid funds. In many other developing countries, the implementation of similar foreign capital policies would have met with international sanctions. Washington tolerated

[5] Krueger (fn. 1), 21; Joungwon Alexander Kim, *Divided Korea: The Politics of Development, 1945–1972* (Cambridge: East Asian Research Center, Harvard University, 1975), 128; Jones and Sakong (fn. 3), 210; David C. Cole and Princeton N. Lyman, *Korean Development: The Interplay of Politics and Economics* (Cambridge: Harvard University Press, 1971).

[6] Krueger (fn. 1), 161; Jones and Sakong (fn. 3), 275–78; Haggard and Moon (fn. 4), 144.

[7] Bank of Korea, *Economic Statistics Yearbook*, 1973.

these policies largely because of the important role Korea has played in supporting what was perceived as a vital American security interest in northeast Asia. Because of this security relationship, Korea has been able to pursue nationalistic economic policies without assuming many of the costs ordinarily associated with them.

After the military coup, institutional changes further reinforced the autonomy of the Korean state. In 1961, major reforms placed all banks and investment decisions under the control of the government.[8] These reforms, established through a combination of incentives and coercion, made all Korean enterprises subservient to the government's guidance as they became dependent on it for funding their survival and expansion. In addition, the government assumed control over trade unions and farmers' cooperatives.[9] When these conditions combined to form a politically insulated state, governmental elites were able to conduct foreign economic policy independent of the interests of any particular societal group.

Although the existence of a politically autonomous state does not assure any particular type of economic performance, a high degree of state autonomy may well be a necessary precondition if political institutions and policies designed to control foreign economic penetration are to be developed and consistently maintained throughout a process of industrial development. Not many developing countries possess the political autonomy necessary to implement such policies.

II. Legal and Bureaucratic Mechanisms of Control over the Sectoral Flow of Foreign Capital and Technology[10]

Since the military regime came to power in 1961, the enhancement of national sovereignty, both strategic and economic, has been the primary objective of the Korean government. In the economic sphere, this has meant rapid industrial expansion. To achieve this objective, the state

[8] Koo (fn. 4), 173; Edward Mason et al., *The Economic and Social Modernization of the Republic of Korea* (Cambridge: Harvard University Press, 1980), 336–37.

[9] Koo (fn. 4), 171.

[10] This study draws on several sources of primary data collected by the author while conducting field research in Korea between August 1984 and December 1986, and during the summer of 1988. These data fall into three general categories: (1) statistical yearbooks and other sources of published Korean government data and policy papers; (2) internal government documents made available to the author by government officials who were interviewed for this study (interviews were conducted with approximately 60 government officials who were directly involved in formulating and implementing economic policy); (3) data obtained from approximately 75 foreign business and banking officials in Seoul who agreed to complete questionnaires and to be interviewed.

pursued a strategy to develop key "strategic" industries. It established powerful institutions, such as the Economic Planning Board, to design and implement a series of five-year plans that sequentially targeted sectoral development. With regard to the inflow of foreign capital, these plans were highly detailed in identifying specific projects and technologies explicitly delineating the sectors and projects where domestic producers possessed the necessary technical know-how as well as those for which foreign technical expertise was necessary. Foreign capital and technical inflows were coordinated with development plans. Projects for which domestic technical knowledge was adequate were closed to foreign penetration, while those for which foreign technical assistance was necessary were designated as open to foreign penetration. These plans and the designated "closed" and "open" sectors were strictly adhered to in the implementation of industrial development because the Korean leadership feared that financially strong and technologically advanced foreign firms would undermine the development of domestic firms and the ability of the state to control key economic and political decisions.

The government therefore established elaborate legal and bureaucratic mechanisms designed to guide foreign capital inflows into a pattern that would enhance the implementation of development plans while it did not hinder the advancement of domestic enterprises.[11] Foreign investment was considered to benefit the Korean national interest only when it produced in a manner that complemented official planning—principally by developing targeted import substitution industries that were technologically too complex for domestic producers, or by producing exports that earned foreign exchange. Foreign investment was not considered beneficial to national objectives when it competed with domestic firms on the Korean market.

Control over the inflow of foreign capital and technology was legally established under the Foreign Capital Inducement Act (FCIA) of 1961 and its subsequent revisions. The FCIA became the principal legislation to control the inflow of foreign capital loans, foreign direct investment, and foreign technology. It was designed to increase the attractiveness of lending to and investing in Korea while placing the control of the pattern of inflows in government hands.

Under this act, the government guaranteed the repayment of all foreign capital loans flowing into Korea; these required the approval of the Economic Planning Board (EPB) which, in turn, was contingent on their

[11] Bohn-young Koo, *Industrial Structure and Foreign Investment: A Case Study of Their Interrelationship in Korea* (Seoul: Korean Development Institute, Working Paper No. 8402, 1984), 7.

flowing to industries targeted by state development plans. This screening process allowed the state to control the size and sectoral direction of foreign loans entering the Korean economy.[12]

The FCIA is also the principal legal document by which the government induces and regulates direct foreign investment and foreign technological inflows. This act guarantees the right of foreign firms to repatriate profits and offers generous tax and other financial incentives to foreign investment in "open" sectors that have been approved by the EPB.[13] The approval process is also tightly controlled under the act, and only those foreign inflows perceived to be complementary to the state's industrial development strategy are allowed to enter the Korean economy.[14]

All foreign investment proposals submitted under the FCIA are required to specify the nature of the project, the level and source of investment capital, the specific technology that will be transferred, the ratio of capital to labor, the methods of training and promoting local personnel, the projected levels of output, the ratio of output to exports, and the ratio of foreign to domestic equity holdings. By requiring a well-specified investment proposal, the government can calculate the impact of that in-

[12] In an interview, Bohn-young Koo, Special Economic Advisor to the Minister of the Economic Planning Board, explained the loan approval procedure. He emphasized that, in order to approve a foreign loan, the government must be convinced that a high level of certainty exists that the borrowed capital will generate a rate of return in excess of interest payments; that it will aid the balance of payments; that it will flow to a targeted industry; and that it will flow to a producer with sufficient technical and marketing expertise to meet projected output levels.

Koo's argument was strongly supported by my 1986 survey of 18 major American, Japanese, British, and French bank officers in Seoul. All of them agreed that the screening process allows the state to control the size and sectoral flow of all foreign capital loans. They further agreed that the screening process assures a high level of certainty that foreign loans will aid the balance of payments. Moreover, all but two of the bank officers in the survey agreed that the screening process assures a high level of certainty that borrowed funds will flow to targeted industrial development.

One bank officer who had served in Brazil for five years in the 1970s asserted that in Brazil, the government had little idea of where foreign loans would flow or whether they would lead to increased production. Often, loans were not monitored and were used for consumption. In Korea, the state monitors foreign loan funds closely to ensure that they flow only into approved investment and that output projections are met after the production facility becomes operational.

[13] Dong Sung Cho, "Incentives and Restraints: Government Regulation of Direct Investment between Korea and the United States," in Karl Moskowitz, ed., *From Patron to Partner* (Cambridge: Harvard University Press, 1985), 45; Ministry of Finance, *Investment Guide to Korea* (Seoul: Ministry of Finance, 1984), 19; Larry Westphal et al., *Korean Industrial Competence: Where It Came From* (World Bank Staff, Working Paper No. 469, 1977), 18.

In 1981, the making of approval decisions on foreign capital and technological inflows was transferred to the Ministry of Finance. For a comparison of incentives to foreign investment offered by various governments in Asia, see "Special Survey of Foreign Investment in Asia," *The Economist*, June 23, 1979, pp. 8–9.

[14] Koo (fn. 11), 18, 26–28.

vestment on the Korean economy and increase the state's capacity to monitor and control the foreign investor's activity after arrival.

To implement and enforce the FCIA and to regulate interaction with the international economy in general, the state developed elaborate bureaucratic mechanisms. The principal agency in this regard has been the Foreign Capital Deliberation Committee which was first attached to the EPB and then transferred to the Ministry of Finance in 1981. The membership of this committee is made up of high-ranking officials from the various economic ministries. All foreign loan applications, investment proposals, and technological transfer agreements must be approved by this committee. Every proposal for foreign economic inflows is individually passed from the Foreign Capital Deliberation Committee to the ministries concerned with the subject matter. It is then reviewed in detail by each ministry to assess its economic and technological impact on the Korean economy as well as its political desirability.[15] All of the ministries must approve each individual foreign inflow before the Foreign Capital Deliberation Committee can approve it. If the Ministry of Finance considers a proposal to be unfavorable to the balance of payments, if the Ministry of Trade and Industry considers it to be in competition with a domestic producer, or if the Ministry of Science and Technology rules that the proposal calls for the importation of technology already available on the domestic market, the application is denied.

The fact that a sector or project is listed on the FCIA as open to foreign investment does not guarantee that a proposal in that area will be approved by the government. The FCIA merely identifies general sectors that are open for proposed foreign activity; it provides guidelines for those foreign inflows that the government may desire and legal prohibitions on those it does not. The highest level of the bureaucracy maintains the authority and capacity to rule on a specific foreign inflow based upon its analysis of the impact of that inflow on the Korean economy.

If more than one foreign investor is interested in investing in a particular sector or project, the government will use investment proposals as bids and will bargain with the competing foreign investors to make the terms of their proposals more advantageous to Korean economic interests. For example, the Dupont Corporation proposed to build a $150 million production facility to produce TIO_2, a whitening agent used in many products manufactured in Korea. (The domestic chemical industry does not possess the capacity to produce TIO_2.) According to the FCIA, a foreign investor can legally hold 100 percent of equity in a TIO_2 project;

[15] Cho (fn. 13), 46-47.

that is what Dupont proposed to do. While the Foreign Capital Deliberation Committee was reviewing Dupont's proposal, Kerr McGee submitted a plan to build a similar facility as a joint venture with a Korean firm. The committee, without officially rejecting Dupont's proposal, notified Dupont of the Kerr McGee project and suggested that Dupont also find a joint venture partner. According to the general manager of the Dupont operation in Seoul, the only way Dupont will get approval, regardless of the FCIA, is to outbid Kerr McGee by offering a greater degree of technological transfer and by requesting fewer financial incentives. Dupont has since resubmitted its proposal as a joint venture with a Korean partner.

Thus, the legal and bureaucratic mechanisms are directly intended to manage foreign economic penetration into the Korean economy so as to assure that it flows to sectors that fit the country's development planning. The objective is to control and shape foreign capital inflows into a sectoral pattern that complements the broader economic and political objectives of the state rather than the specific interests of foreign capital.[16]

III. Decoupling Foreign Penetration to Enhance the Expansion of Domestic Enterprise

In its attempt to steer foreign inflows into a pattern that would facilitate the growth of domestic enterprises, the Korean government has adopted a policy of "decoupling" foreign capital and foreign technology from foreign direct investment. It attempts to deliver necessary components to domestic producers rather than to allow foreign investors to provide the entire package of production.

Korean leaders have consistently viewed uncontrolled foreign direct investment as highly threatening to national development goals due to the foreign economic and political influence it tends to engender and to the expropriation of capital that occurs over the long term.[17] Although Korean economic planners realized that foreign capital and technology would be necessary to facilitate rapid industrial expansion, they actively restricted foreign direct investment and assisted domestic firms to acquire necessary capital and technology wherever a targeted sector could be developed without foreign ownership.

[16] T. Jefferson Coolidge, Jr., "The Realities of Korean Foreign Investment Policy," *Asian Affairs* 12 (July/August 1981), 370–85; Dal-joong Chang, *Economic Control and Political Authoritarianism: The Role of Japanese Corporations in Korean Politics 1965–1979* (Seoul: Sogang University Press, 1985), 174–78.

[17] Moskowitz (fn. 13), 11.

Korea has tended to view foreign capital loans as a more desirable form of foreign capital penetration than direct investment.[18] For both political-strategic and economic considerations, Korean economic planners considered long-term concessional loans in particular as more desirable than direct investment. Loan funds could be channeled to domestic producers by the state's financial institutions. Korean ownership would reduce dependence on foreign producers and (at least in principle) assist producers who were more tied to national security and national survival than foreign concerns would be.[19] As long as borrowed funds generated productive outcomes, loans were perceived to be less costly to the national economy than direct foreign investment. Loans do produce an immediate drain on the national economy, but when they are paid off, domestic ownership and further profits remain in the country. By contrast, foreign direct investment continually drains the economy through repatriation of profits and transfer pricing. "From a long-term point of view, the burden of foreign exchange outflow is greater on equity investment than on loans."[20]

As a result, foreign direct investment has played a minor role in Korean capital formation, and loans have played a much more significant

[18] In an interview, Bohn-young Koo stated that "loans have always been preferred over direct investment because more benefits stay in Korea. Direct foreign investment is restricted to those projects where foreign marketing or technical skills are required. If only capital was required, the government would attempt to borrow it. After the 1974 oil shock, when funds became readily available, this strategy could be followed relatively easily."

In a survey of foreign bank officers in Seoul (see fn. 10), 81% of respondents stated that the government has consistently followed a strategy of encouraging loans for foreign capital needs.

[19] This conclusion is based upon a series of interviews and questionnaires (see fn. 10). According to the International Monetary Fund's representative in Seoul, "the Korean foreign capital accumulation strategy has been to borrow capital and license technology wherever possible in order to avoid foreign equity ownership." The general manager of the Chase Manhattan Bank in Seoul asserted that "it has been the conscious and managed policy of the Korean government to use external borrowing to promote growth of the economy. At the same time, it is the conscious and managed policy of the Korean government to ensure that domestic industry is not foreign-controlled. That means more debt and less foreign equity."

The general manager of Dupont in Seoul maintained that "if technology is locally available, a direct foreign investment will simply not be allowed"; a senior official with Westinghouse stated, "the Koreans would prefer to operate totally without foreign investment if that would be possible."

According to 92% of the foreign bank officers surveyed, Korean borrowing to finance foreign capital requirements has allowed a greater degree of national control over major economic decisions than a similar level of foreign direct investment would have.

[20] *Study on the Criteria for Foreign Direct Investment and Joint Venture in Korea* (Seoul: International Management Institute, Korea University, 1972), 54–56. In interviews, this argument was also made by Kwan-tae Shin, Deputy Director, Investment Promotion Division of the Korean Ministry of Finance; by Bohn-young Koo, Special Economic Advisor to the Minister of the Economic Planning Board; and by Dong-gyu Shin, Assistant Director of the Foreign Capital Policy Division of the Ministry of Finance. Also see Sung-hwan Jo, *Direct Foreign Private Investment in South Korea: An Economic Survey* (Seoul: Korean Development Institute, Working Paper No. 7707, 1977), 68.

. Between 1962 and 1971, direct foreign investment constituted only rcent of net inflow of foreign capital; between 1972 and 1976, 7.9 ent; and between 1976 and 1984, less than 4 percent.[21] Comparison other industrializing countries shows that direct foreign investment ween 1967 and 1971 constituted only 3.7 percent of net foreign capital flow in Korea, while the comparable figure was 33.8 percent in Brazil, 6.6 percent in Mexico, and 26.1 percent in Thailand. Between 1972 and 1976, a period when import substitution engendered several large capital-intensive foreign investments in Korea, the ratio of direct foreign investment to net foreign capital inflow was still only 7.9 percent, as compared to 22.9 percent in Brazil, 16 percent in Mexico, and 28 percent in Thailand.[22] The reason for this difference is that Korea's development strategy has limited foreign direct investment strictly if its principal contribution was merely a source of capital, and has encouraged foreign loans that could be channeled to Korean producers.

In addition, any necessary foreign technology has preferably been adopted through licensing agreements. This policy, again, is intended to facilitate the development of domestic production and to reduce the level of penetration of foreign direct investment.[23] In targeted sectors, where domestic technology was inadequate to meet planned production requirements, Korea preferred to license technology and to pay the necessary licensing fees rather than to allow direct foreign investment. The Ministry of Science and Technology has developed bureaus and "think tanks" whose staffs spend the majority of their time attempting to locate foreign technological sources willing to lease technology to Korean producers. With the right combination of loans and access to technology, rapid economic expansion could be accompanied by maximum levels of domestic ownership and control over the production apparatus.

Foreign direct investment was preferred and sought after only if it could provide a technology that was necessary to develop a targeted sector, and if that technology was not available domestically or through a technological license. The government offered attractive economic incentives when a particular technology was deemed crucial to the development of a targeted sector, and when it was available only through foreign investment. Economic incentives were also available to export producers, but other types of investments were restricted. By these means, the Korean state was able to shape the pattern of its interaction

[21] *Major Statistics of the Korean Economy* (Seoul: Economic Planning Board, 1985).

[22] Economic Analysis and Projections Department: International Bank of Reconstruction and Development, *Consolidated Balance of Payments* (Washington, DC: World Bank, May 19, 1978).

[23] Moskowitz (fn. 13), 11.

with the international economy and to facilitate national ownership of the means of production.

IV. The Empirical Pattern of Foreign Direct Investment

The inflow of foreign direct investment into Korea has been channeled to sectors designated for development and into projects where Korean producers lacked independent technical capacity.[24] *In the early stages of industrial development, for example, foreign investment was channeled into the development of synthetic fibers for textile production and into major import substitution projects in petroleum refining and chemical fertilizers. By 1968, these sectors contained 75 percent of all foreign investment in Korea. Once these import substitution investments* were in place, the government designated textiles and electronic consumer goods as sectors for export. In turn, these exports were intended to finance the development of capital-intensive chemical and heavy industrial sectors. Throughout the 1970s, that is where foreign investment was principally channeled. By 1980, foreign investment was concentrated in chemicals (25 percent), machinery (16.7 percent), electronics (12.2 percent), metal products (4.8 percent), and construction (4.5 percent).

Because the state has attempted to minimize the level of foreign investment by limiting its entrance into only those sectors and projects that were considered necessary for and compatible with its development objectives, the size of foreign direct investment equity holdings in Korea is minimal compared to total stock or to total foreign capital usage.[25] Between 1962 (when the first foreign investment was accepted) and 1985, only $2.65 billion of foreign investment was approved by the government, and only approximately $1.9 billion actually arrived.[26] Foreign direct investment has been effectively excluded from most sectors that produce consumer goods. Unlike most developing countries, it has also been effectively excluded from agricultural and other primary production sectors. (See Table 1.) Restrictions on foreign firms owning land and on

[24] Charles A. Barone, "Dependency, Marxist Theory, and Salvaging the Idea of Capitalism in South Korea," *Review of Radical Political Economics* 15, no. 1 (1983), 48. In a survey of 45 foreign investors in Korea (fn. 10), 100% of the respondents asserted that the Korean state exercised much more control over the pattern of direct foreign investment than other capitalist developing states with which they were familiar. All respondents also maintained that the Korean government effectively excluded foreign investment in areas where domestic producers could meet production demands.

[25] Koo (fn. 11), 26–28.

[26] Ministry of Finance data (fn. 10). For the same period, comparable figures are: almost $8 billion in foreign direct investment in Taiwan, $17 billion in Mexico, $24 billion in Brazil, and $7 billion in Singapore (U.S. Department of Commerce, Overseas Business Reports, 1987).

TABLE I
FOREIGN INVESTMENT BY ECONOMIC SECTOR
(NUMBER OF PROJECTS)

	1962–66	*1967–71*	*1972–76*	*1977–81*	*1982–85*	*Total*	*Percent*
Mining and Agriculture	1	8	26	15	16	66	5.6
Manufacturing	14	139	391	149	283	976	82.3
Foodstuffs	1	4	3	5	28	46	3.9
Textiles	3	15	44	5	6	73	6.2
Wood products	—	—	2	1	—	3	0.3
Chemicals	—	20	56	24	36	136	11.5
Medical products	3	4	3	4	28	42	3.5
Fertilizer	2	—	2	—	—	4	0.3
Petroleum	1	2	1	—	1	5	0.4
Ceramics	—	7	12	3	7	29	2.4
Metals	—	13	33	17	20	83	7.0
Machinery	1	19	62	37	51	170	14.3
Electronics	3	27	121	21	66	238	20.0
Transportation	—	1	4	2	2	9	0.7
Other	1	26	48	25	38	138	11.6
Social Overhead Capital	—	17	33	33	61	144	12.1
Finance	—	2	3	5	7	17	1.4
Construction	—	7	9	11	35	62	5.2
Electricity	—	—	—	2	—	2	0.2
Transportation	—	5	4	4	2	15	1.3
Hotels and tourism	—	3	17	11	17	48	4.0
Total	16	163	450	197	360	1,186	100.0

SOURCE: Ministry of Finance documents (Seoul, December 31, 1985); based upon approval of investment applications.

those investing in agriculture, mining, or consumer goods productior
strictly enforced.

Of the total value of foreign investment approved since 1962, 59.2 p
cent is in manufacturing and 40 percent is in service-related sector
(See Table 2.) Manufacturing investments have been heavily conce
trated in chemicals, electronics, and machinery; together they accou
for 39.3 percent of total foreign direct investment, or 66.3 percent of tot
manufacturing foreign investment. These three industries, combine
with textiles, transportation equipment, metal products, medical proc
ucts, and food processing, constitute 53 percent of total foreign inves
ment, or 89 percent of all manufacturing foreign investment. All of ther
have been designated as target sectors by state planners. Foreign invest
ments in these areas have played a significant role in Korean industrial
ization by providing foreign-exchange-earning exports or technologie
for import substitution that have created necessary linkages for domesti
industrial producers.

In service sectors of the economy, foreign direct investment has beer
concentrated in three principal areas. Hotels and tourism have led al
others in attracting foreign investment (over $793 million, or 29.9 per-
cent, of the total). The Korean government has viewed tourism as a rel-
atively costless way of earning foreign exchange. In the fifth five-year
plan (1982–1986), the tourism industry was targeted for development,
but Korea had little experience in operating world-class hotels and tour-
ist attractions. Principally led by Japanese firms owned by families with
Korean ancestry, the tourist industry has been the sector of the Korean
economy most open to foreign investment, attracting over 43 percent of
foreign investment approved by the government.

During the past ten years, finance and construction have also received
large shares of foreign investment. Fearful of foreign dominance of fi-
nancial markets, the Korean government had severely restricted foreign
investment in the financial sector until the late 1970s. As the economy
grew to be more technologically advanced and diversified, however, it
became apparent that financial services in Korea were antiquated. Mod-
ern technology was needed to increase the efficiency of capital accumu-
lation and resource allocation. During the fourth five-year plan (1976–
1982), the government began to feel confident in its ability to control the
financial sector and no longer feared foreign investment as a major threat
to national control. With the objective of gaining access to advanced fi-
nancial technologies, it encouraged leading international banking cor-

[47] Ministry of Finance data (fn. 10).

Foreign Investment by Economic Sector
(value in U.S. $1,000)

	1962–66	*1967–71*	*1972–76*	*1977–81*	*1982–85*	*Total*	*Percent*
Mining and Agriculture	102	943	6,672	7,162	8,809	23,688	0.9
Manufacturing	22,897	59,925	425,937	392,468	669,204	1,570,431	59.2
Foodstuffs	100	1,056	1,729	29,276	38,628	70,789	2.7
Textiles	759	4,048	57,505	2,562	7,675	72,549	2.7
Wood products	—	—	263	850	—	1,113	0.1
Chemicals	—	11,626	146,143	120,258	100,391	378,418	14.3
Medical products	313	2,654	2,082	9,377	62,613	77,039	2.9
Fertilizer	20,500	—	21,325	150	—	41,975	1.6
Petroleum	—	7,845	15,407	8,283	5,000	36,535	1.4
Ceramics	—	5,649	4,379	4,231	14,232	28,491	1.1
Metals	—	8,975	30,003	21,318	12,352	72,648	2.7
Machinery	20	5,287	30,330	44,139	197,757	277,533	10.5
Electronics	901	10,614	82,995	107,765	183,894	386,169	14.5
Transportation	—	40	25,542	19,485	21,499	66,566	2.5
Other	304	2,131	8,234	24,774	25,163	60,606	2.3
Social Overhead Capital	—	11,805	132,630	188,113	728,300	1,060,848	40.0
Finance	—	1,621	11,340	52,472	55,830	121,263	4.6
Construction	—	3,307	15,856	37,344	54,661	111,168	4.2
Electricity	—	—	—	3,395	—	3,395	0.1
Transportation	—	783	3,412	24,437	3,355	31,987	1.2
Hotels and tourism	—	6,094	102,022	70,465	614,454	793,035	29.9
Total	22,999	72,673	565,239	587,743	1,406,313	2,654,967	

Source: Ministry of Finance documents (Seoul, December 31, 1985); based upon approval of investment applications.

porations to invest in Korea. In addition, joint ventures with K
banks were promoted through financial incentives. By 1985, the fin
sector accounted for 4.6 percent of total foreign investment approv.

Prior to the late 1970s, the government had also restricted foreig
vestment in the construction sector. Domestic construction was prot
for firms such as Hyundai, Ssangyong, and Dong A. After the oil (
however, a unique overseas market opportunity presented itself t
Korean construction industry; in order to take advantage of it, p
toward foreign investment in the construction sector needed to be
tered.

Utilizing petrodollars, several Middle Eastern countries embarked
large-scale development projects in the 1970s. Because construction w
opened to international bidding, Korean firms (which had reputations
efficient producers who met completion deadlines and provided rela
tively inexpensive labor) frequently were in a position to underbid
American, European, and Japanese concerns. The expansion of Korean construction firms to the Middle East offered an opportunity to earn several billion dollars in foreign exchange. However, several of these Middle Eastern contracts required advanced construction technologies that the Korean firms did not possess. To acquire these technologies, the government offered very attractive financial incentives to foreign construction firms that would enter into joint venture agreements and provide specific construction techniques to Korean firms doing business in the Middle East. By 1985, foreign investment in the construction sector constituted 4.2 percent of all foreign investment approvals. Over 82 percent of these investments arrived after 1977; the majority provided complex technology related to overseas construction projects.

The empirical record of the pattern of foreign direct investment is closely related to the state's industrial planning and policy. Strict control over the inflow of foreign direct investment has allowed the Korean economy to utilize desired foreign technical expertise while shaping the pattern of foreign investments in a manner that protects and nurtures the development of domestic enterprise.

V. Strategies to Utilize Foreign Direct Investment to Enhance the Expansion of Domestic Enterprise

In addition to controlling the sectoral flow and decoupling the principal components of foreign economic penetration, the Korean government has effectively pursued policies designed to limit the activity of foreign direct investment in the Korean economy and to increase the

tic producers' benefits. All foreign investments entering Korea negotiate elaborate investment agreements with the government, attempts to specify the conditions under which the foreign inves- y operate. These conditions may include (1) a joint venture to be d with a Korean partner, through which the latter will have or btain, in a specific period of time, financial and operational control he enterprise; (2) an agreement as to the levels of capital invest- , output, and export, the level and type of technological transfer, the ision of raw materials, and the access to foreign markets that the ign investor will provide; and (3) an agreement on the divestiture of eign-held equity to the domestic partner at a specified future date.

The effectiveness of these strategies varies by sector and by case; it pends on the government's perception of how necessary a particular eign investor is to the achievement of development objectives, and on w attractive the Korean market is to the investor. If the government's conditions are too restrictive and the investment can be placed elsewhere, the investment will not be placed in Korea. If the investment involves a critical technology, Korean development objectives may therefore be eopardized by requiring excessively stringent conditions. In such cases, he government will have to reduce the conditions placed upon the for- ign investor in order to obtain a critical investment. On the other hand, f the foreign investor perceives the Korean market to be lucrative, the overnment will attempt to tighten the conditions in the investment greement. Natually, if there is more than one interested foreign inves- or, the government's position is enhanced. In such cases, the foreign ivestors may be required to make competitive bids; the government nay then accept the bid that maximizes its position.

Central to the issue of benefits to the domestic economy is the share of urplus extracted from the domestic economy. Profit remittance, transfer ricing, royalty payments, and the division of local and export markets re issues directly related to how surplus is divided between a foreign ivestor and the domestic economy. One primary strategy for increasing enefits to the Korean economy has been the government's requirement nat, whenever possible, foreign investments form joint ventures with Korean firms as a condition of entry. Even in those sectors listed in the oreign Capital Inducement Act as open to foreign direct investment, ureaucratic control mechanisms maintain the principle that a minimum f 50 percent of equity be held by a local partner. The only foreign in- estments that are approved with foreign ownership levels exceeding 50 ercent are those that will produce exclusively for the export market or nat will transfer a technology designated as necessary for the achieve-

ment of planned development objectives. The rationale for this policy is that joint ventures will increase transfers of technological and managerial skill to domestic firms and facilitate domestic capital accumulation to a greater degree than would fully foreign-owned investments. In a joint venture, the Korean partner could, over time, learn the technological and production processes. He would then be able to lead any expansion that may be required in the future, as well as adapt his newly acquired technical knowledge to related operations. The policy that foreigners cannot constitute a proportion of the board of directors larger than that of their share of stock ownership further permits greater Korean control over a joint venture's decision-making process than would occur within a wholly foreign-owned venture.

Although the government's policy of promoting joint ventures has affected the pattern of foreign ownership in a manner favorable to domestic producers, foreign investment has not been encouraged by it. Obviously, multinational corporations generally do not desire to form joint ventures and to give up control over their operations. In order for a foreign investor to be willing to participate in a joint venture, he needs to be assured of at least a minimum level of profitability. Incentives such as guaranteed access to the domestic market and the sole right to produce in a given sector for a specific period of time can attract some foreign investors. In Korea's case, this has been particularly true in labor-intensive sectors and in those sectors that require less sophisticated technology. To attract more sophisticated technology, however, the government has had to permit larger shares of foreign ownership or to guarantee profit levels. Foreign producers of export merchandise have also been granted permission to hold larger shares. This is reflected in the pattern of foreign ownership levels over time and by sector.

In 1968, 18.6 percent of firms with foreign financial participation were totally foreign-owned, while 24.8 percent were majority foreign-owned.[28] Most of them were small firms engaged in export production of textiles and electronic consumer goods. Those that produced exclusively for export markets were granted majority or total foreign ownership. Foreign firms that produced for the domestic market were required to form joint ventures and to hold less than majority ownership. At that time, the two largest were an oil refinery and a fertilizer production facility. Both were major targeted import substitution projects. Because the government wanted to have Korean involvement, however, it proposed a joint venture and guaranteed a minimum profit rate of 15 percent, a five-year tax

[28] Economic Planning Board data (fn. 10).

holiday, exclusive rights to the domestic market, and exclusive rights to supply raw materials. The Korean partner was granted 50 percent ownership.

In order to encourage foreign investors, particularly Japanese, to establish production facilities in the chemical, machinery, textile, and electronic sectors, the Korean government started to ease its restrictions on the level of foreign ownership in these sectors in 1968. By 1972, 21.6 percent of all foreign-invested firms were totally foreign-owned. At that time, however, the government, conscious of Japanese ownership of production facilities in Korea, began to realize that little technological transfer to Korean producers occurred where a firm was 100 percent foreign-owned.

In 1973, the Foreign Capital Inducement Act was changed to limit majority ownership in foreign investments to firms that brought in government-specified technology for targeted import substitution or that exported over 80 percent of output.[29] By 1980, the proportion of wholly foreign-owned firms had declined to 14 percent, and only 12.2 percent were majority foreign-owned. By 1985, only 13.3 percent of all foreign investments were wholly foreign-owned.[30]

That is an extremely low figure by international standards. In a study of sixty-six countries conducted in 1977, Korea ranked lowest in the proportion of wholly foreign-owned firms—approximately 17 percent. The second-lowest was Israel with 30 percent; Japan was third with 33 percent.[31] By 1985, only 13.3 percent of all foreign-invested firms in Korea were wholly foreign-owned. Between 1981 and 1985, only 10.7 percent of the total of 401 foreign investment proposals approved by the government were for wholly foreign-owned firms; majority foreign ownership was granted to 28.1 percent.[32] Korea has thus effectively pursued a highly restrictive strategy toward foreign ownership share.

Once a foreign firm enters into a joint venture with a Korean firm, technical and managerial expertise begins to be transferred. As agreed upon in the terms of the joint venture, the foreign partner is required to train the Korean partner's personnel in managerial, technical, engineering, accounting, and other related production techniques and skills. Several foreign managers in Korea related that a joint venture with a Korean firm involves a high level of personnel rotation. Korean personnel

[29] Koo (fn. 11), 21.

[30] Ministry of Finance data (fn. 10).

[31] Joan P. Curhan, William H. Davidson, and Rajan Suri, *Tracing the Multinationals* (Cambridge: Cambridge University Press, 1977), 314, as cited in Bohn-young Koo, *The Role of Foreign Direct Investment in Korea's Recent Economic Growth* (Seoul: Korean Development Institute, 1983), 26.

[32] Ministry of Finance data (fn. 10).

assigned to joint venture projects are often transferred to other subsid-
iaries of the Korean partner after they have acquired the necessary
knowledge to perform their tasks. New personnel must then be trained
by the foreign partner while the original employees apply their knowl-
edge to other production processes of the Korean firm. After the Korean
personnel has learned the technology necessary to operate the joint ven-
ture enterprise, expansion in that sector is carried out by the domestic
partner exclusively, and that sector is closed to further foreign invest-
ment. Eventually, production thus becomes dominated by Korean pro-
ducers.

Joint ventures also increase the likelihood that the profits generated by a foreign-invested firm will be reinvested in the Korean economy. If a developing country directly restricts the repatriation of profits, it will have difficulty in attracting desired foreign investment. For this reason, the Foreign Capital Inducement Act explicitly guarantees the right of foreign investors to repatriate profits. Some of the surplus that the Korean government would prefer to utilize in the form of investment capital thus goes abroad. However, since profits must be shared by the foreign investor's Korean joint venture partner, a portion is reinvested in the Korean economy; moreover, the overall level of dividends paid out may be reduced because of the government's incentives to the domestic partner to reinvest a large share of earnings in the joint venture's operation.

During interviews with foreign investors and with officials of the American and the Japanese Chambers of Commerce in Korea, a common complaint surfaced concerning the conflict over the generation of profits and the issuance of dividends between foreign and domestic joint venture partners. The state provides low-cost financing to Korean firms in targeted sectors, largely based upon a firm's level of output and/or export expansion. If a Korean firm does not increase output or export levels according to state quotas, it risks losing access to future subsidized loans and other economic privileges. For this reason, Korean firms are not as concerned with profits and dividends as they are concerned with expansion. In a joint venture, there is therefore a continual struggle over the proportion of reinvestment to dividends paid. A Westinghouse official in Seoul pointed out that, if the local partner refuses to issue a dividend, remittance of profit is difficult. This conflicting goal orientation between domestic and foreign partners has limited the dividends as well as the repatriation of foreign capital, thereby increasing the level of gross domestic investment over that which would have occurred if the project had been wholly foreign-owned.

The petroleum-refining industry serves as a good example of how int venture operations have led to increased Korean technological now-how and the "nationalization" of an industry. In 1962, Korean tate planners designated oil refining as one of the first strategic import-ubstitution projects. However, no Korean firm possessed the needed technology. The Gulf Oil Corporation was approached to develop a refining operation in Korea, but it was reluctant because of the government's insistence on a joint venture relationship. In order to persuade Gulf to enter into a joint venture with the Korean Oil Company, the government offered it highly favorable terms including tax concessions, ree land, total financial and operations control, monopoly rights on the upply of crude oil to the operation, and a guaranteed minimum profit evel of 15 percent. Gulf's original investment amounted to $4.8 million n 1963. By 1973, Gulf had remitted over $10.5 million in profits, excludig charges for the transportation and sale of crude oil to its Korean peration.[33] The partnership served as an import-substitution project nat would aid the balance of payments and guarantee the supply of a eeded raw material; perhaps more important to long-range development objectives, it also offered the opportunity for Korean engineers, echnicians, and managers to learn the oil-refining process.

As the Korean economy grew, the need for refined petroleum inreased. Gulf offered to expand its capacity. However, the Korean govrnment wanted to reduce its dependence on only one refining operation nd on only one supplier of crude by diversifying foreign sources. By iat time, economic growth had made the Korean economy more attracve to foreign investors and allowed the government to open up bidding n the second refinery. In 1965, a joint venture was formed between altex and the government-selected Gold Star Corporation to establish orea's second oil refinery. Caltex, itself a joint venture of Standard Oil f California and Texaco, received generous incentives including price ipports, tax concessions, and financial and operational control.

The demand for refined petroleum continued to grow along with the orean economy. In 1968, Union Oil Corporation entered into a joint nture with Korean Explosives Corporation. In 1976, after the first oil iock, in an attempt to extend its suppliers to Middle East oil sources, ie Korean government persuaded the National Iran Oil Corporation to rm a fourth joint refining venture, with Ssangyong Corporation of Koa. In the first two ventures, Gulf and Caltex had been granted operaonal and financial control under the terms of the investment agreement

[33] Chang (fn. 16), 137.

and were required to supply all technological know-how. In 1968, the Korean state granted operational control to Union Oil, but only 50 percent decision-making power in financial matters; it did not grant guaranteed profits to Union. By 1976, Korea had gained the necessary technical expertise to operate a refining operation. Under the terms of the National Iran Oil Corporation–Ssangyong venture, the Korean partner held total operational and 50 percent of financial control.

In the late 1970s, the government began to nationalize foreign-held equity in petroleum-refining joint ventures. When, after the fall of the Shah, the new Iranian government agreed to sell its equity to Ssangyong, the Korean government provided financing to Ssangyong to purchase the Iranian corporation's share. After the second oil shock in 1979, several oil companies found themselves in short supply of crude and in need of restructuring their worldwide portfolio. Gulf wanted to sell its Korean operation not only because of crude oil shortages, but also because of legal problems in the United States that linked the chairman of Gulf Oil to a bribery scheme with Korean President Park's Democratic-Republican Party.[34] Union Oil needed to restructure its worldwide portfolio; it wanted to sell its interest in Korea partly because of problems in repatriating dividends that resulted from Korean Explosives' actions to reinvest profits in corporate expansion.[35] Thus, both Gulf and Union were willing to sell their equity shares at relatively low prices. Because Korean partners now possessed the necessary technological skills to operate the refining facilities, foreign participation was no longer needed. The government encouraged the domestic partners to purchase the foreign interest by extending subsidized financing. By 1983, the only foreign-invested firm in the Korean oil-refining sector was Caltex—and it was required to renegotiate its original investment agreement so as to grant operational and financial control to Gold Star. A fifth oil-refining operation was established in 1986 as a joint venture between British Petroleum and Kukdong Corporation of Korea. British Petroleum holds 40 percent of equity and supplies the raw materials for the venture; Kukdong holds financial and operational control.

Today, Korea refines 100 percent of its domestic requirements, purchases its crude oil on a contractual basis or on the international spot market, hauls the bulk of it in its own ships, and holds total financial and operational control over its petroleum-refining sector.

[34] Ibid., 134–36.

[35] Information on the Korean repurchase of foreign-held equity in the oil refinery sector has principally been obtained in interviews with Caltex officials, Korean government officials, and professors at Seoul National University School of Business and Management.

Another strategy employed by the Korean government in its attempt to control and increase the benefits the domestic economy receives is to demand that investment agreements specify the factors of production and other services that foreign investors will provide to the Korean economy. The state thus enhances its capacity to control their activity and to require them to perform and not to exceed a specific role.[36]

In targeted import-substitution sectors, the Korean government has considered technological transfers and raw material supplies to be the primary contribution of foreign investors. It has offered generous incentives to induce their entry, but explicitly states in the investment agreements the specific technological transfers and the level of raw material to be supplied. For example, Amoco was offered generous tax concessions and a guaranteed profit level for five years when it entered into a joint venture with Samsung Corporation of Korea in 1974 to manufacture P.T.A., a primary ingredient for producing polyester fiber, plastic containers, and cassette tape. In this major import-substitution project for the Korean textile, electronic, and other related industries, the terms of the investment agreement specified the level of all raw materials Amoco would supply and the method of technical transfer to Samsung.[37]

In labor-intensive sectors, where the technological gap between foreign and domestic firms was relatively small, the Korean government has allowed only export production by foreign investors. This policy was motivated by the rationale that foreign firms with greater capital resources and marketing experience might take domestic market shares away from Korean producers. The only way in which labor-intensive foreign investments would aid the domestic economy was through the foreign investors' knowledge of, and access to, foreign markets. Accordingly, the investment agreement required them to export a specific portion of their output and restricted their domestic sales. After 1973, most labor-intensive foreign investments were licensed only if they exported 100 percent of their output.[38] By 1978, 39 percent of foreign-invested firms were required to export 100 percent of their output. Over 25 percent of the firms that were permitted to sell on the domestic market had to sell at least 50 percent of their output abroad. Ceilings on output levels further minimized the domestic market share that some of these firms could control. Fewer than 40 percent of foreign investments had no con-

[36] Koo (fn. 11), 82.

[37] This information was collected in interviews with Amoco and Samsung officials in Seoul.

[38] Eul Young Park, "An Analysis of the Trade Behavior of American and Japanese Manufacturing Firms in Korea," in Moskowitz (fn. 13), 24, 31.

dition regarding export levels, and these investments were highly concentrated in targeted import-substitution sectors.[39]

Specific terms of entrance requirements are normal for gaining access to the Korean economy. In 1986, 38 percent of agreements with foreign investors stated specific export-level requirements, 80 percent stated specific technological transfer requirements, 36 percent stated specific requirements for raw material supply, and 28 percent of the investors agreed to help their Korean partners gain access to export markets. At least one of the above requirements was specified in the investment agreements of over 92 percent of all foreign firms; at least two or more were specified in over 68 percent.[40]

In order to protect and facilitate domestic ownership and production, the Korean state did not approve a foreign investment if it would compete with a domestic producer. As a parallel to this policy, no foreign firm was allowed to take over or purchase an existing Korean firm.[41] If a domestic producer was technically capable of production in a particular area of a sector that needed to expand, the state would attempt to formulate capital from domestic or international sources and allocate it to the domestic producer. If a domestic concern wanted to sell its operation, the state required the purchaser to be a Korean national. In the course of research for this study, I could document only one case of a Korean firm purchased by a foreign investor. That was in the 1960s, when A-Dong electric, an export-oriented electronic assembly operation, was purchased by a Japanese firm. After this, the Economic Planning Board stopped all purchases of domestic firms by foreign interests. This is in sharp contrast to Raymond Vernon's estimate that one-third of all foreign subsidiaries in developing countries were formed through the purchase of domestic firms, and to discussions such as Peter Evans's account of the "denationalization" of domestic firms in Brazil.[42]

The Korean government has pursued a policy of actively encouraging the Korean repurchase of equity held by foreign firms. Of $1.3 billion in foreign direct investment between 1962 and 1983, Korean firms repurchased $263.5 million by 1983.[43] Over 20 percent of foreign direct invest-

[39] Westphal et al. (fn. 13), 53.

[40] Survey of 45 foreign investors (fn. 10).

[41] *Business Korea*, July 1985, 76.

[42] Vernon, *The Storm over the Multinationals: The Real Issues* (Cambridge: Harvard University Press, 1977), 72; Evans (fn. 2, pp. 103–25) defines denationalization as the displacement of local enterprises by foreign ones. He traces the denationalization of several industries in Brazil and highlights a Brazilian industrial ownership structure where, because of the superior capital and technological levels of multinational corporations, local enterprise tends to survive only in small-scale, non-capital-intensive enterprises.

[43] Economic Planning Board data (fn. 10).

ments were reacquired after Korean nationals had gained technical knowledge of the production process. The government encouraged these purchases by including specific divestiture requirements in investment agreements, and by allocating subsidized loans for the purchase of foreign-held equity.

In the labor-intensive sectors, where its bargaining position is strongest vis-à-vis foreign investment, the Korean government not only requires foreign investors to enter into a joint venture with a Korean partner; they must also agree to those terms of the investment agreement that require their divestiture of equity. Under this provision, the Korean partners have the option to purchase a certain equity level (usually 15 percent, or enough to give them a majority share) after a certain period (usually ten years) or when the foreign partners have repatriated a specific percentage (usually 150 percent) of the original investment.[44] At the time of the stock repurchase, operational and financial control is of course transferred to the Korean partners.

In the late 1960s and early 1970s, several Japanese manufacturers of electronic consumer goods and of textiles established production facilities in Korea to take advantage of low labor costs. Under the terms of entry, the Korean government granted tax privileges, low utility costs, and other incentives in conjunction with financial and operational control. However, it also required joint ventures with Korean partners as well as divestiture agreements. For the Korean economy, these foreign investments increased levels of export and employment and created the opportunity to expand domestic control after local producers had gained technological and other necessary skills.

Many of the triggering mechanisms of the divestiture agreements were reached by the mid-1970s. The government usually supported the Korean partners in exercising the equity purchase options by providing low-cost financing. In many cases, the foreign companies were opposed to giving up their majority positions, but the Korean government pressured them to abide by the original agreement. Several foreign companies disposed of all of their equity rather than become minority partners. Mashushita, Toshiba, Sony, and others sold off their ventures in Korea and relocated in Singapore, the Philippines, or elsewhere. Some did not leave Korea, but reduced their equity levels. Teijin, with a 50 percent interest in the Korean firm Sunkyong (a producer of polyester fiber), was forced in 1976 to reduce its share of equity to 33.5 percent and to give up

[44] The information on the divestiture of foreign-held equity was obtained through an interview with the Seoul branch manager of a major Japanese bank who was personally involved in the negotiations to buy out several Japanese firms.

financial and operational control. Toray Corporation had to reduce its share of a joint venture with Korean Polyester from 40 to 19.8 percent, and Mitsui's share in the same firm declined from 10 to 8.2 percent; the Korean partner thus gained majority ownership and control.[45]

By means of this strategy, Korean corporations gained ownership and control of several electronic consumer-goods and textile companies. The local market was then protected for these producers, allowing them exclusive rights at home; subsequently, they also began to compete successfully in international markets with many of the same companies that had been their joint venture partners.

The Korean government has established several means to ensure the compliance of foreign-invested enterprises with the appropriate laws and with the terms of their investment agreements. The Ministry of Finance, the Ministry of Trade and Industry, the Ministry of Science and Technology, the Ministry of Agriculture, the Ministry of Health, the Customs Bureau, and the National Tax Service all have agencies that regularly monitor various aspects of the activities of foreign investors.

The Ministry of Finance is responsible for monitoring foreign exchange matters. A predominant argument in favor of foreign direct investment in developing countries is that it provides a source for needed capital. Several studies have shown, however, that a large share of the investment capital of foreign-owned subsidiaries is actually formulated in domestic capital markets,[46] thereby defeating the theory. Because the Ministry of Finance has control over financial institutions and the allocation of investment capital, it has been able to guard against foreign investors using domestic funds for investment. Of the foreign investors who responded to the survey conducted for this study in 1986, 86 percent declared that all their investment capital had been formulated abroad. All of the remaining 14 percent had borrowed funds from foreign bank branches in Seoul, using the parent corporations to guarantee the loan. Thus, these funds did not truly originate with domestic Korean sources. The restrictions ensure that domestic savings are utilized for the development of domestic enterprises.

The Foreign Investment Control Bureau of the Ministry of Trade and

[45] Chang (fn. 16), 192.

[46] See Peter Evans, "National Autonomy and Economic Development: Critical Perspectives on Multinational Corporations in Poor Countries," in Robert O. Keohane and Joseph S. Nye, Jr., eds., *Transnational Relations and World Politics* (Cambridge: Harvard University Press, 1972), 328–30; Richard Barnet and Ronald Muller, *Global Reach* (New York: Simon & Schuster, 1974), 152–65; Economic Commission for Latin America, *External Financing in Latin America* (New York: United Nations, 1965); David T. Devlin and George R. Kruer, "The International Investment Position of the United States," *Survey of Current Business* 50, no. 1 (October 1970).

Industry is reponsible for monitoring the operational activities of foreign-invested firms. Foreign investors must submit detailed reports on a monthly basis to the bureau on such matters as output levels, import and export activities, profit rates, domestic content, taxes paid, and personnel affairs. The bureau then uses these reports to assure that investors comply with their investment agreements.

The Bureau of Customs also monitors imports and exports; among other things, it is responsible for ensuring that only those capital goods directly related to approved production processes, and not produced in Korea, are imported by foreign firms. This policy is designed to limit the extent of transfer pricing by foreign firms and to increase the use of Korean-produced capital goods in foreign-invested production processes.

One hundred percent of respondents surveyed for this study asserted that they had been monitored by various state agencies on a regular basis to ensure their compliance with Korean law and with the terms of their investment agreements. Several foreign investors recalled instances in which the government determined that a foreign investor was not in compliance with local law or with its investment agreement. In these cases, the government would levy fines or revoke privileges offered in the investment agreement, such as tax reductions or low rates on utilities, until the investor complied with the agreement.

Conclusion

In Korea, a nationalistic state, acting in a relatively autonomous manner has shaped the pattern of foreign capital integration to facilitate the expansion of domestic firms. As a consequence, industrial development is predominantly owned and effectively controlled by Korean nationals. In a broader theoretical context, this demonstrates that the nation-state in a developing country is capable of exercising a high degree of control in establishing the pattern of its linkages with the international economy. Development strategy and governmental mechanisms can, to a large extent, shape foreign economic inflows into a sectoral and ownership pattern that facilitates both industrial development and domestic entrepreneurial expansion. The ability of the state to act in such a manner is a variable not adequately addressed or specified in neoclassical or dependent development literature.

Though the capacity and the will to do so vary greatly from one state to another, national governments can and do exercise important controls over their economic development. To a significant degree, they can determine the conditions under which access to their markets will be

granted. The principal criterion concerning the relationship between a developing economy and the international economy should not simply be whether foreign captial is beneficial to the host country's economic development (because it provides capital, technology, managerial skills, and market expertise) or whether it is detrimental (because it increases dependency levels, utilizes inappropriate technology, expropriates national surplus, and facilitates foreign control over key economic decisions). The focus should be on the structures and strategies that can be developed by the state in order to induce necessary inputs in a manner that will maximize the benefits to the domestic economy. The state is the only institution that can advance comprehensive strategies to shape the pattern of foreign capital to fit the needs of the national economy.

Review Article

UNDERSTANDING INTERNATIONAL TRADE POLICIES

An Emerging Synthesis

By JOHN S. ODELL*

Robert E. Baldwin, *The Political Economy of U.S. Import Policy*. Cambridge: The MIT Press, 1985, 238 pp.
Jagdish Bhagwati, *Protectionism*. Cambridge: The MIT Press, 1988, 147 pp.
I. M. Destler, *American Trade Politics: System under Stress*. Washington, DC and New York: Institute for International Economics and the Twentieth Century Fund, 1986, 366 pp.
David A. Lake, *Power, Protectionism, and Free Trade: International Sources of U.S. Commercial Strategy, 1887–1939*. Ithaca, N.Y.: Cornell University Press, 1988, 242 pp.
Helen V. Milner, *Resisting Protectionism: Global Industries and the Politics of International Trade*. Princeton: Princeton University Press, 1988, 329 pp.
Robert M. Stern, ed., *U.S. Trade Policies in a Changing World Economy*. Cambridge: The MIT Press, 1987, 437 pp.

WHY do governments act as they do in matters of international trade? During the 1980s, international trade problems multiplied and gripped public attention in many countries. Although an exensive literature already existed, scholarship on trade policies and poliics expanded in proportion to the problem: following a trend that began n the 1970s, many economists and political scientists turned their tradiional tools to the subject. As a result, we now have a large literature that oheres around a common question, but remains, or appears to remain, lisparate in the theoretical answers it defends and the research methods t employs. Fragmentation of political-economic knowledge is not news; s we know, the two disciplines parted company during the nineteenth entury. Even today, much of economics shows no signs of contact with

* The author appreciates helpful comments by Jonathan Aronson, Thomas Biersteker, Benjamin J. Cohen, Peter Katzenstein, and Thomas Willett. None of these friends is responsible for any errors or omissions. The author also thanks the Ford Foundation and the Social cience Research Council for support for related research.

political science, and vice versa. The good news is that a better, unified theory—combining elements of economics, political science, and history—is gradually emerging in the area of international trade policy and elsewhere.

In this essay, I shall single out four prominent and distinct theoretical themes woven into scholarship of the mid- to late 1980s, deliberately passing over other contributions. The four perspectives emphasize market conditions, leaders' values and beliefs, national political institutions, and global political-economic structures. Analysts working in particular traditions generally have well-known intellectual and professional reasons for presenting their ideas boldly and for claiming as much as possible for them relative to their competitors; that tendency is visible in some of the volumes under review. I shall maintain, however, that this fashionable tendency, while stimulating in the short term, is not adequate as a long-term strategy for our disciplines.

Not surprisingly, each of the four influential perspectives proves to have some theoretical value, yet none has managed to dominate the field. Nor will any one be adequate by itself as a framework within which to unify knowledge about trade policies. Most of the shortcomings and contributions of each approach will be apparent from this review. These books, however, together with other research under way, also reveal how a creative synthesis is evolving and may develop further. My main argument will be that analysts are increasingly resisting the temptation to defend bold but relatively narrow traditional ideas as sufficient, and are explicitly blending valuable insights from multiple perspectives in the same study—and that more of them should do so. After presenting evidence for this reading, I shall conclude with special attention to places where the fabric needs to be strengthened in the future.

The volumes under review share several features. The guiding orientation is more explanatory than prescriptive, though several also generate advice for dealing with current problems. They focus on explaining national policies that affect imports, though export promotion does appear occasionally. The central questions are located primarily at the national rather than the international or regime level of analysis. The studies tend to concentrate on the United States, introducing other industrial democracies only for comparative purposes or not at all.

In several respects, the books under review also illustrate the diversity of this literature. They differ as to their dependent variables and in their theoretical visions. Their methods represent quantitative and nonquantitative techniques as well as static interindustry contrasts and longitudinal designs.

These volumes constitute only one subset of recent work on trade policy formation and politics. Other interesting segments investigate the concept and practice of strategic trade policy,[1] trade sanctions and East-West trade,[2] Japan's economic performance, policies, and power,[3] and various international trade conflicts, negotiations, and institutions.[4]

The Market Perspective

The first perspective argues generally that market conditions shape trade politics and policies. The most familiar interpretation in the United States (though not in all countries) runs as follows: free trade is best for the nation as a whole, but certain national industries are weak in the market; they organize politically to press for protection, and government provides protection for their products. Household consumers who reap the gains from liberal trade usually do not express their interests politically. E. E. Schattschneider still provides the standard description of pressure groups run amok during enactment of the infamous Smoot-Hawley tariff of 1930.[5]

Neither Schattschneider nor much of the subsequent research on pressure groups attempted, however, to provide a rigorous theory that would predict which sectoral markets would generate protection, or to identify

[1] See parts of Robert Stern's volume not reviewed here, and Paul R. Krugman, ed., *Strategic Trade Policy and the New International Economics* (Cambridge: The MIT Press, 1986). This literature is surveyed by J. David Richardson, "The Political Economy of Strategic Trade Policy," *International Organization* 44 (Winter 1990), 107–35.

[2] David Baldwin, *Economic Statecraft* (Princeton: Princeton University Press, 1986); Gary Clyde Hufbauer and Jeffrey J. Schott, assisted by Kimberly Ann Elliott, *Economic Sanctions Reconsidered: History and Current Policy* (Washington, DC: Institute for International Economics, 1985); Bruce W. Jentleson, *Pipeline Politics. The Complex Political Economy of East-West Energy Trade* (Ithaca, NY: Cornell University Press, 1986).

[3] Bela Balassa and Marcus Noland, *Japan in the World Economy* (Washington, DC: Institute for International Economics, 1988); Kent E. Calder, "Japanese Foreign Economic Policy Formation: Explaining the Reactive State," *World Politics* 40 (July 1988), 517–41; Takashi Inoguchi and Daniel I. Okimoto, eds., *The Political Economy of Japan*, vol. 2, *The Changing International Context* (Stanford, CA: Stanford University Press, 1988); Clyde V. Prestowitz, Jr., *Trading Places: How We Allowed Japan to Take the Lead* (New York: Basic Books, 1988); Richard J. Samuels, *The Business of the Japanese State. Energy Markets in Comparative and Historical Perspective* (Ithaca, NY: Cornell University Press, 1987); Karel van Wolferen, *The Enigma of Japanese Power: People and Politics in a Stateless Nation* (New York: Knopf, 1989).

[4] See Vinod K. Aggarwal, *Liberal Protectionism: The International Politics of Organized Textile Trade* (Berkeley: University of California Press, 1985); John A. C. Conybeare, *Trade Wars: The Theory and Practice of International Commercial Rivalry* (New York: Columbia University Press, 1987); Jock A. Finlayson and Mark W. Zacher, *Managing International Markets: Developing Countries and the Commodity Trade Regime* (New York: Columbia University Press, 1988); and Gilbert R. Winham, *International Trade and the Tokyo Round Negotiation* (Princeton: Princeton University Press, 1986). In addition, Benjamin J. Cohen reviews two of the present books and others in "The Political Economy of International Trade," *International Organization* 44 (Spring 1990), 261–81.

[5] Schattschneider, *Politics, Pressures, and the Tariff* (New York: Prentice-Hall, 1935).

the general conditions most likely to lead governments to provide it, reject demands for it, or change its levels over time. During the 1970s, some scholars began to construct such a theory, grounded in neoclassical rational choice or public choice premises, and to test it systematically. Public choice analysts, typically abstracting from much detail, depict politics as a process of exchange among individuals who behave according to short-run personal self-interest—exclusively so in the purest variants. These treatments in fact postulate a market for protection itself, with rent-seeking industries on the demand side trading their votes and campaign contributions to politicians on the supply side. Governments, too, are seen as rational utility maximizers, sometimes as unitary actors and at other times as sets of self-interested individual politicians and officials responding to the "political market" and maximizing under institutional constraints. Interest groups are still prominent players, but the analysis penetrates beneath their activities to identify market conditions that will drive them. Thus when markets shift, so will the distribution of political pressures and consequently government policies as well, moving the political economy to a new equilibrium.

One recent contribution from this perspective is the essay "Endogenous Protection in the United States, 1900–1984," by Stephen P. Magee and Leslie Young (Stern, 145–95). This piece begins with a standard neoclassical Heckscher-Ohlin-Samuelson open economy and adds lobbies, political parties, and voters. The still highly parsimonious model is then presented as a partial explanation for U.S. import policy for the entire twentieth century. The authors explain:

> By endogenous protection, we mean that trade restrictions are set by maximizing, self-interested behavior by all of the economic agents, lobbies, political parties, and voters. . . . The concept of endogenous policy generates the following paradox: in a highly competitive political system, the equilibrium policies are not under the control of policymakers. In the case of protectionism, trade restrictions balance the power of narrow interests (lobbies) against that of broad interests (voters). The trade policies are equilibrating variables that clear political markets; they are analogous to prices in product markets, which balance the quantity demanded against the quantity supplied. The policymaker is like an auctioneer in a product market. . . . Empirically long-run protection should be explained by those exogenous variables that drive the behavior of groups . . . that favor and oppose protection. . . . Our empirical work indicates that approximately two-thirds of the changes in U.S. tariffs this century are explained by economic variables suggested by an endogenous tariff model (unemployment, inflation, and the U.S. terms of trade). (pp. 145–46)

The core model of Robert Baldwin's *The Political Economy of U.S. Impo[r]t Policy* also emerges from this tradition; it begins with the premise tha[t]

free trade is optimal, and then introduces imperfections in the political-economic market to account for protective policies. Baldwin, however, goes much further than Magee and Young in adding other complexities.

Industrial Market Conditions

One way in which national markets change over the long term is through structural shifts in their factor endowments. Magee and Young hypothesize that, in the United States, labor benefits disproportionately from protection; as labor's share in production declines, its relative political clout should also decline, as should the fortunes of the political party tied to it—and hence, protection as well.

During the medium or short run, the authors reason, factors will tend to stay out of politics as long as times are good; but, "when any factor's economic fortunes decline, it transfers effort out of economic activity and into lobbying and political activity" (Stern, 161). For instance, as U.S. terms of trade improve, with lower prices in the United States for Japanese cars, American blue-collar jobs and wages are undermined, and the opportunity cost of lobbying declines for labor. The United Auto Workers compensate by lobbying for higher protection, and Washington grants it. By the same token, as inflation accelerates, the pro-export lobby and even ordinary voters will exert greater demands for import liberalization.

Over time and across industries, a market analyst would expect policy preferences to vary at least with the industry's or firm's position in the world economy and with its economic health. More specifically, with rising import penetration and declining returns on capital, wages, and employments rates, industries can be expected to favor increasing protection for themselves, ceteris paribus. On the other hand, the greater an industry's dependence on exports and on foreign investments, the more it should resist demands for protection and favor import liberalization. Another staple hypothesis is that protection is a collective good for the industry seeking it, as is open trade for consumers; therefore, rational free riding will reduce a large group's political pressure below the level justified by its interests.

Empirically, the picture drawn by recent research in this tradition is less than entirely clear, as usual. Studies vary as to their basic methods, including both quantitative and small-n comparative case studies. Different works select different dependent variables to measure trade policy or industry demands as well as different data series, cases, and time periods. Econometric tests show that models are typically incomplete, and case studies naturally have difficulty establishing the generality of their findings.

We do find evidence confirming several basic ideas, however. In the United States, tariff rates certainly did decline between 1900 and 1984 while the labor/capital ratio was falling, though a great many other things were also changing. Magee and Young report the expected positive effects on tariff rates for the lagged terms of trade and current unemployment, and the expected negative effect for current inflation. They also test the model on the number of antidumping petitions filed in recent years.

At the same time, there are good reasons to be skeptical of the assumption that the key divide falls between labor and capital. Workers and investors at Apple Computer have common trade interests that differ from those of both groups at Bethlehem Steel; classes often do not behave as units in trade politics. The Magee-Young model is also confronted with the paradoxical evidence that Republican presidential administrations were associated with greater protection than those led by Democrats, the party of labor. The authors explain that this is because Republican macroeconomic policies generated more unemployment and lower inflation, which in turn activated greater pressures for protection at the sector level and quieted anti-protection forces. Democratic macroeconomic policies tended toward the opposite conbination of inflation and unemployment, dampening pro-protection micropressures. In short, "protection is a byproduct of each administration's macropolicy" (Stern, 87).

Baldwin provides original econometric tests of several models which assume that government is simply an intermediary, responding to voters and groups attempting to maximize short-run direct self-interest. He supplements the tests with rich case studies. His book disaggregates U.S. trade policy into the independent decisions that arise in the contemporary institutional setting. He observes the votes of individual members of Congress on trade legislation, votes of commissioners of the U.S. International Trade Commission (ITC) on industry petitions for special import relief, presidents' decisions on ITC recommendations, and the varying tariff reductions offered by presidents on different products in multilateral trade negotiations in the 1960s and 1970s.

Congressional voting on the 1974 trade act is found to be "consistent with the hypothesis that Congress is responsive to the competitive problems of import-sensitive industries and labor unions" (p. 77). Higher import penetration and rising import penetration are also associated with smaller tariff cuts offered in negotiations (chapter 4), but import penetration does not appear to be a particularly strong influence on the ITC's decisions to recommend special relief from import competition, nor on

presidential decisions in such cases (pp. 103–14, 129–33).[6] Baldwin also produces significant evidence that industries which are relatively dependent on exports and on foreign investment tend to have lower degrees of import protection and deeper tariff cuts (chapter 4).

Contrary to expectations, however, tests of the collective action hypothesis tend not to support it very strongly. To be sure, household consumers fail to contribute much to political campaigns for free trade, but when producer industries are considered as groups, relations between industry concentration ratios and protection levels or proposed cuts tend to be insignificant or even to have the wrong sign.[7] The reason could be that protection and open trade alike can be made less-public goods by narrowing the definition of the product to be fought over. Within narrower categories, small-number solutions to the collective action problem are more feasible. Moreover, economic statistics do not capture the presence of private institutions. Some trade interests are represented by industry associations formed and supported for a variety of purposes, while other trade interests are not so represented. Firms sometimes look beyond earnings in the short term to overcome problems of collective action.

Interdependence and Anti-Protection Pressurf

Over the long term, U.S. companies have become more dependent on exporting and foreign investment; according to several recent studies, this is a central influence restraining contemporary U.S. trade policy. The new literature is adding more careful attention to the other side of he domestic political struggle, and is looking beyond household con-umers. One of these books, Helen Milner's *Resisting Protectionism*, be-gins with a historical puzzle: why the 1920s ended with Smoot-Hawley and trade war, but the 1970s did not—despite several similarities be-ween the periods. She uses a longitudinal comparison to raise doubts about the importance of the much-debated hegemony hypothesis, noting hat the relative international economic position of the United States was quite similar at the ends of the two periods. Passing over other political actors, Milner emphasizes long-term changes in market conditions

[6] The sample for the latter tests is already limited to cases with relatively high import ompetition; this truncation might account for departures from common sense.

[7] See also Réal P. Lavergne, *The Political Economy of U.S. Tariffs: An Empirical Analysis* Toronto: Academic Press, 1983), Table 7.2; William R. Cline, *Exports of Manufactures from Developing Countries* (Washington, DC: The Brookings Institution, 1984), Table 2-6, on non-ariff barriers; and Thomas A. Pugel and Ingo Walter, "U.S. Corporate Interests and the Political Economy of Trade Policy," *Review of Economics and Statistics* 67 (August 1985), 471, on company activity toward trade legislation.

(other than the labor/capital ratio) and rational self-interested calculation by U.S. companies.

Milner's research posits industry's political demands, not policy itself, as the dependent variable, and thus addresses directly only part of the general question that organizes this literature. On the other hand, unlike many studies coming from this first perspective, the research on sectoral demands is more direct, using industry's congressional testimony and other evidence of political behavior rather than relying on common but less adequate economic proxies. Milner selects the six U.S. industries that faced the greatest increases in import penetration in the 1920s and another six from the 1970s, and conducts a detailed study of each.

She concludes that the firms that were least dependent on both exports and multinational operations (type I industries) tended to respond to import competition by calling for protection in both periods. Those that exported more but did not have multinational operations (type II) were found to have more divided policy preferences. Industries more internationalized in both senses (type III) tended to prefer free trade in both periods, despite strong import penetration at home (chapter 6). Furthermore, her research indicates that increasing interdependence over the decades increases the likelihood that firms in the same industry will diverge in the degree of their international orientation, dividing more sectors politically when they confront an import challenge. All in all, the demand for protection was more muted in the 1970s than in the 1920s. Moreover, if Milner's result holds for industries facing heavy import competition, a liberal policy stand seems even more likely for internationalized sectors not facing such pressure. Although the book is based on case studies, they are "least likely" cases for her hypothesis. Methodologically, this volume provides a model for selecting a small number of cases for maximum analytical gain.

Overt opposition to campaigns for product-specific protection in the U.S. expanded several-fold after the period examined by Milner. The producers of sugar, textiles, and automobiles increasingly found themselves counter-lobbied by other special interests that depended on imported components, by retailers who had staked their competitive strategies on access to foreign goods, by exporters, and in some cases by multinational companies. One study concludes that although the opposition did not clearly win all fights, it did make a difference in the 1980s, helping pro-trade politicians and officials to restrain the degree of new protection or, in some cases, to deny the requests altogether. This countermobilization was much less in evidence, however, during consider-

ation of general (not product-specific) legislation that could make administrative criteria more restrictive.[8]

MACROMARKET CONDITIONS

Not surprisingly, the 1980s also caused American analysts to think more about macroeconomic influences on trade policy. The dramatic expansion of U.S. fiscal and external deficits and the rocketing of real interest rates and the dollar into the stratosphere made these influences difficult to ignore. One of the oldest macroeconomic notions had held that bad times favor protection, and vice versa. New research has provided better confirmation of this business cycle hypothesis.[9]

In "Macroeconomics and Protection," Rudiger Dornbusch and Jeffrey A. Frankel (Stern, 77–130) suggest several additional connections. Their basic logic links changes in the macroeconomy to changes in exchange rates and trade flows, which in turn affect output and employment aggregates at home; these lead, finally, to shifts in political pressures for import barriers. Dornbusch and Frankel remind us that it is the country's *relative* macroeconomic position that counts for the trade balance. They document the well-known paradox that when a country recovers faster than its trading partners, it sucks in new imports faster than it boosts exports, thereby worsening its trade balance (Stern, 85). A declining aggregate trade balance may be favorable for an increase in the average level of protection, insofar as it draws politicians' attention to trade rather than to other issues, and undermines their confidence in a liberal policy.

Dornbusch and Frankel's favorite argument is the exchange-rate hypothesis. The dollar's rise made exports less competitive and imports more so. The import flood touched off more pro-protection pressure, the argument goes, and may also have undermined political pressure against protection by U.S. export groups such as farmers. The authors conclude from empirical data that the rising dollar of the early 1980s was a major cause of the plummeting American trade balance, and that a failing currency will improve it (Stern, 89). This point became controversial in po-

[8] See I. M. Destler and John S. Odell, *Anti-Protection: Changing Forces in U.S. Trade Politics* (Washington, DC: Institute for International Economics, 1987), as well as Bhagwati, *Protectionism*. Little work has been done to investigate such political opposition in other countries.

[9] See Wendy E. Takacs, "Pressures for Protectionism: An Empirical Analysis," *Economic Inquiry* 19 (October 1981), 687–93; Guilio M. Gallarotti, "Toward a Business Cycle Model of Tariffs," *International Organization* 39 (Winter 1985), 155–87; and James Cassing, Timothy McKeown, and Jack Ochs, "The Political Economy of the Tariff Cycle," *American Political Science Review* 80 (September 1986), 843–62.

litical debates of the later 1980s, when the dollar did fall sharply, but the trade deficit did not decline as rapidly or by as much as many had hoped it would. Dornbusch and Frankel also find that high real interest rates reduced employment in the U.S. automobile and textile/apparel industries, both of which gained additional protective measures during the decade.

This first, market-based perspective contributes essential insights; no one would deny that market conditions are relevant for a government's economic policies. More particularly, the public choice tradition has improved the rigor and generality of our knowledge of which sectoral or macromarket conditions are most strongly related to which U.S. policy variables (though empirical results are not fully consistent). By adopting some simplifying assumptions, the analyst is able to employ deductive models to explain phenomena that fit within those limits, producing a more satisfying certainty and precision than is found in studies that view the issue from most other perspectives.[10] We also now have greater evidence that pro-trade interests express themselves through politics.

This theory is inadequate and even misleading if taken narrowly or alone, however. It is one thing to focus attention on economic markets; it is another to suppose that politics itself is a market, and especially that government is nothing more than an auctioneer responding to society. Many political specialists would dismiss this extreme vision out of hand. In any case, treating politics as analogous to a market of individuals obscures the society's institutionalized structures of authority and their effects. By assuming individuals as primary and their preferences as given—an essential component that yields advantages—the perspective also acquires a blind side, namely, the extent to which authority structures shape the very definition of individual interests or preferences and the formation of groups, as well as the behavior of both in particular situations.

Devotion to a narrow version of public choice analysis by some may be a product of excessive concentration on the more pluralist circumstances of North America and similar regions. In principle, a utility function can hold anything we wish to insert; the notion could be stretched to model a dictator who cares about unemployment, for example. But it has not been shown that we gain by conceiving of trade policies in authoritarian or totalitarian systems as products of choices by publics-cum-auctioneers, relative to a top-down view starting with elites pure and

[10] For a general review and case for the affirmative, see Bruno S. Frey, "The Public Choice View of International Political Economy," *International Organization* 38 (Winter 1984), 199–223.

imple. Even in the United States, presidents, the Department of Agriculture, and congressional committees are not well understood as auctioneers or traders; they have authority over their "exchange partners." They make decisions that can put a company out of business, and sometimes do so, but no industry, even a large one like steel, can destroy a president or a political party. To base conclusions on the premise that their relationship is a free, uncoerced exchange would be misleading, even in the United States.

Moreover, a narrow public-choice interpretation simply omits too much that other research shows to be important. The perspective misses variables internal to the domestic political process that do affect the level of protection. Most of these works also abstract from international relations altogether. While reading many studies coming from a market perspective, we forget the power imbalances, trade wars, cartel arrangements, international organizations, and liberalizing negotiations that have dotted commercial history. These omissions are usually accepted as part of a theoretical strategy favoring parsimony and allowing for complexities only one step at a time. This trade-off between parsimony and validity is familiar, but there may be better ways to finesse it.

The Cognitive Perspective

A second theoretical theme woven into recent thinking about trade policy begins with the premise that political behavior is partly a function of leaders' and publics' values, policy beliefs, and ideologies, and that differences and changes in these ideas will shift policies accordingly. This general argument, springing from roots in Weber and elsewhere, assumes that policy ideas, while affected by material interests, are not simply reducible to them; values and beliefs have more complex origins and can have independent effects on policy content.[11] This second perspective aspires not simply to describe policy more accurately, but to help explain it and its changes.

[11] In one famous passage Weber put his own formulation as follows: "Not ideas, but material and ideal interests directly govern man's conduct. Yet very frequently the 'world images' that have been created by 'ideas' have, like switchmen, determined the tracks along which action has been pushed by the dynamic of interests." *Gesammelte Aufsaetze zur Religionssoziologie*, vol. 1, p. 252, quoted in *From Max Weber: Essays in Sociology*, trans., ed., and with an introduction by H. H. Gerth and C. Wright Mills (New York: Oxford University Press, 1946), 63–64. Albert O. Hirschman contends that the traditional economic approach, which disregarding values such as public morality, presents too simple an account of even such fundamental processes as production and consumption. "Against Parsimony: Three Easy Ways of Complicating Some Categories of Economic Discourse," *Economics and Philosophy* 1 (April 1895), 7–21.

The cognitive perspective has penetrated less into trade studies, so far, than into macroeconomic and other areas, but interest is spreading.[12] Charles Kindleberger, contemplating the remarkable spread of free-trade policies in mid-nineteenth-century Europe, concludes that the most satisfactory explanation is that "Europe as a whole was motivated by ideological considerations rather than economic interests. . . . Manchester and the English political economists persuaded Britain, which persuaded Europe—by precept and example."[13] A classic work on U.S. trade politics in the 1950s, *American Business and Public Policy*,[14] concluded that a businessman's stand on protection could not be explained simply by "what he makes"; policy preferences were also affected by the person's general orientation between isolationism and internationalism, as well as by such factors as whether the executive had traveled overseas and whether another industry or politician had attempted to influence his thinking.

Increasingly during the 1980s, rational choice thinkers conceded the need for direct investigation of beliefs and preferences that cannot be reduced simply to material interests. Participants in one project, for example, reported that "the contributors to *Cooperation under Anarchy* did not specifically set out to explore the role of perception in decision making, but the importance of perception has kept asserting itself."[15]

"One broad conclusion" Robert Baldwin reaches "is that the models focusing exclusively on short-run, direct self-interest are insufficient for

[12] See Judith Goldstein, "The Impact of Ideas on Trade Policy: The Origins of U.S. Agricultural and Manufacturing Policies," *International Organization* 43 (Winter 1989), 31–72. For comprehensive reviews, see John S. Odell, *U.S. International Monetary Policy* (Princeton: Princeton University Press, 1982), 58–78 and passim, and Peter A. Hall, ed., *The Political Power of Economic Ideas: Keynesianism across Nations* (Princeton: Princeton University Press, 1989). For recent related applications, see Emanuel Adler, "Ideological 'Guerrillas' and the Quest for Technological Autonomy: Brazil's Domestic Computer Industry," *International Organization* 40 (Summer 1986), 673–706; Paul Egon Rohrlich, "Economic Culture and Foreign Policy: The Cognitive Analysis of Economic Policy Making," *International Organization* 41 (Winter 1987), 61–92; John S. Odell, "From London to Bretton Woods: Sources of Change in Bargaining Strategies and Outcomes," *Journal of Public Policy* 8 (July–December 1988), 287–316, and other articles in the same issue; Ethan B. Kapstein, "Resolving the Regulator's Dilemma: International Coordination of Banking Regulations," *International Organization* 43 (Spring 1989), 323–47; and Peter M. Haas, "Do Regimes Matter? Epistemic Communities and Mediterranean Pollution Control," *International Organization* 43 (Summer 1989), 377–404.

[13] Kindleberger, *Economic Response: Comparative Studies in Trade, Finance and Growth* (Cambridge: Harvard University Press, 1978), 65. See also Joseph A. Schumpeter, *History of Economic Analysis* (New York: Oxford University Press, 1954), 397–98, and Peter Gourevitch, *Politics in Hard Times* (Ithaca, NY: Cornell University Press, 1986).

[14] Raymond Bauer, Ithiel de Sola Pool, and Anthony Dexter, *American Business and Public Policy: The Politics of Foreign Trade* (New York: Aldine 1972).

[15] Robert Axelrod and Robert O. Keohane, "Achieving Cooperation under Anarchy: Strategies and Institutions," in Kenneth A. Oye, ed., *Cooperation under Anarchy* (Princeton: Princeton University Press, 1986), 226–54, at 247, first pub. in *World Politics* 38 (October 1985).

explaining the wide range of behavior patterns observable in the trade policy arena. Models that include behavior based on either long-run self-interest or concern for the welfare of other groups and the state are also necessary to account for the actions of voters and public officials" (p. 174). His broader public choice study insists that "elected officials may not simply be intermediaries whose behavior reflects the wishes of the electorate or some part of it. Instead they may pursue their own public policy goals on many occasions and still retain their elected positions (p. 13)." Some of the industries receiving the greatest protection tend to have a high proportion of low-skilled, low-income workers, a group not known for its well-oiled lobbying operations. Presidents offer smaller tariff cuts for industries with poorer and less skilled workers and slow employment growth. The ITC is more likely to recommend import relief for industries that have suffered declines in profits and employment.[16]

Jagdish Bhagwati picks up this theme in a different way in *Protectionism*, the published version of his 1987 Ohlin lectures. "Interests and ideology interact to shape these underlying [trade policy] trends. My major theme will be that contrapuntal tendencies can be distinguished in interests and in ideology, and that the deck is not stacked in favor of protectionism" (p. xii). The Depression experience converted many countries to a "pro-trade bias" in the postwar period, which facilitated liberalization under the General Agreement on Tariffs and Trade (GATT). A free-trader himself, Bhagwati also cites turn-of-the-century arguments against English "reciprocitarians." To these, he adds post–World War II theoretical amendments calling into question even the infant-industry case for protection (pp. 26–33). Bhagwati also takes encouragement from the resurgence of neoclassical ideas in developing countries in the 1980s. He heaps sarcasm on the contemporary "manufacturing matters" justification for protection (110–14), and he looks toward the future with qualified optimism, based partly on confidence in the power of this ideology in the world, and partly on growing global interdependence of interests. Presumably the spectacular developments in Eastern Europe as the decade's curtain came down cheered these liberal hopes further. Bhagwati's recent lectures clearly illustrate the movement toward blending insights from traditionally alternative perspectives.

It is one thing to argue that interests are not all that matters; it would

[16] Stephen V. Marks and John McArthur extend this analysis specifically to congressional voting on protection for the automobile industry; they find that ideology also influences this behavior independently of constituency interest, at least when the member faces a low opportunity cost of voting his or her beliefs. "Empirical Analyses of the Determinants of Protection: A Survey and Some New Results," in John S. Odell and Thomas D. Willett, eds., *International Trade Policies: Gains from Exchange between Economics and Political Science* (Ann Arbor: University of Michigan Press, 1990).

be quite another to maintain that perceptions are. Occasionally (not in Bhagwati's book), one does find such claims—namely, that individuals' and society's interests are always seen through particular cultural lenses and are socially constructed; sometimes this borders on suggesting that interests have no objective basis at all. Such an exclusive variant of a cognitive perspective, which assumes that interests are wholly malleable, could be sustained only by ignoring or rejecting much research that documents regular patterns of behavior that correspond to interests objectively defined, as has been discussed above. A depiction leaving structures of interest and institutions invisible would exaggerate the room for maneuver actually available to policy makers. This second approach, too, would be inadequate and misleading if taken alone.

Even when they are considered to be supplementary, such arguments often suffer from a lack of generality and precision—a price the author is often willing to pay in return for richer insight. Some applications, however, assert the efficacy of ideas and ideologies without attempting to refute alternative interpretations. Many recent studies do identify beliefs and values that made a difference to policy in a single case or two, and some do refute or credit alternatives but do not formulate general propositions that would help to explain why some ideas become more influential than others, or which ideas are related predictably to which aspects of policy. The sampling and measurement of beliefs and ideologies often leaves much to be desired (including in some of my own work), and the importance of ideas relative to other influences is often not specified precisely.

The Institutional Perspective

A third thread woven through this scholarship highlights the constraints that national political institutions impose on individuals and groups, and the effects of institutional change on the broad patterns of trade policy. This perspective also boasts a long lineage, and has in fact experienced a revival in recent years. Precisely in reaction against interest-group liberalism and interdependence thinking, some theorists have been attempting to bring the state back in.[17] In its strongest variant, this

[17] Peter B. Evans, Dietrich Rueschemeyer, and Theda Skocpol, eds., *Bringing the State Back In* (Cambridge: Cambridge University Press, 1985). More relevant for trade policy are Peter Katzenstein, ed., *Between Power and Plenty* (Madison: University of Wisconsin Press, 1978); Robert A. Pastor, *Congress and the Politics of U.S. Foreign Economic Policy* (Berkeley: University of California Press, 1980); G. John Ikenberry, David A. Lake, and Michael Mastanduno, eds., *The State and American Foreign Economic Policy* (Ithaca, NY: Cornell University Press, 1988); and David Mares, "Domestic Institutions and Shifts in Trade and Development Policy: Colombia 1951–1968," in Odell and Willett (fn. 16), 193–224.

perspective begins by assuming that institutions rather than individuals come first. Institutions establish and defend the particular order and initial distribution of property rights under which market exchange takes place. Individuals' very conceptions of their places and potentials in this social order are learned partly from within their particular historical and institutional environment, so that economic liberalization can have fundamentally different effects in different settings. At a minimum, the political behavior of individuals and industries responds to the constraints and opportunities presented by these institutions. Over the long run, national political institutions may crumble or shift in response to market conditions, new ideas, or international structural change—but the long run may take quite some time.

Peter A. Hall constructs a relatively strong version of this argument, primarily for explaining the contrasting paths of macroeconomic and industrial rather than trade policies in Britain and France.[18] Hall also contributes to the synthesis by featuring effects of policy ideas and using some public-choice insights. Most of the books under review also mention dimensions of U.S. political institutions other than parties. Baldwin, for example, emphasizes the differences between trade policy formation within the Congress, in the International Trade Commission, and in the White House, as well as the processes of mutual influence among these bodies.

Baldwin and particularly I. M. Destler in *American Trade Politics: System under Stress* spotlight a change in tariff-setting institutions during the 1930s crisis as pivotal for the subsequent liberalizing trajectory and creation of the GATT. The long-term institutional argument, which is supplementary to Magee's factor endowments, Milner's industry globalization, and international hegemony, centers on Congress's voluntary delegation of authority to the president in the 1934 Reciprocal Trade Agreements Act. Until then, Congress had voted on the specific tariff for each industry; subsequently, the president had the authority to reduce tariffs if he could negotiate reciprocal cuts abroad.[19] This shift created what Destler insightfully calls "protection for Congress"—greater protection, that is, from pressure to do harmful things that Congress knew it could not resist doing if pressed. Of course, politicians had no intention of silencing themselves, but they preferred to be free to make great noises

[18] Hall, *Governing the Economy: The Politics of State Intervention in Britain and France* (New York: Oxford University Press, 1986).

[19] Also see Stephan Haggard, "The Institutional Foundations of Hegemony: Explaining the Reciprocal Trade Agreements Act of 1934," *International Organization* 42 (Winter 1988), 91–120, which links this innovation to requirements of international trade negotiations, and thus to the fourth analytical perspective.

on behalf of their constituents while knowing that the president would make the most consequential decisions and provide them with political cover. They also delegated authority only for a few years at a time, thereby creating periodic opportunities to hold the president to account for reciprocity. Thus, even if market conditions and group rent-seeking had continued unchanged after 1934, special interests faced a more formidable barrier to political success than before. Meanwhile, strong congressional committees were able to kill protective bills introduced by a particular industry before they got to the floor of the full House or Senate. During the Kennedy administration, the creation of a new agency, the Special Trade Representative (now the U.S. Trade Representative), added another institutional counterweight for managing organized pressures.

During the 1970s and 1980s, however, this system came under greater stress and eroded seriously, as Destler shows quite effectively. The opening of the economy—the system's achievement—itself permitted greater local stress in the form of import penetration, which generated intensified demands from well-known industries. At the same time, the institutional democratization of Congress made management more difficult. Strong committees became weaker, less able to keep protective bills off the floor where members would have to take a public stand. The Reagan White House was also less attentive than it might have been to providing political leadership to manage the trade-politics vacuum.

To strengthen this system, a less sanguine Destler proposes improved macroeconomic policy, better microeconomic adjustment, and organizational restructuring. In particular, he would enhance the influence of the U.S. Trade Representative in the White House by slimming that office down in size and abolishing its rival, the Department of Commerce.

In a related line of thinking, Judith Goldstein maintains that particular American policy ideas became institutionalized in law and bureaucracy at different stages during the twentieth century, including the 1930s, and thereby continue to exert influence long after other ideas have become more popular.[20] She concentrates on the roles of the president and the International Trade Commission in enforcing these laws. Goldstein, Destler, and others thus fuse the second and third perspectives.

This institutional approach implies that trade policy formation under different organizational and legal arrangements will yield corresponding differences in policy content. The books under review do not provide

[20] Goldstein, "Ideas, Institutions, and American Trade Policy," *International Organization* 42 (Winter 1988), 179–218.

much cross-national evidence with which to explore this issue, though of course much has been written about Japan in particular (see fn. 3). France, too, is often contrasted sharply with the United States, but according to Milner, the institutional difference there is less than we have been told. She conducts additional case studies of French industries facing import competition in the 1970s, and she finds that her U.S. results are mostly confirmed by the trip across the Atlantic. In France, as in the United States, "firms' growing ties to the world economy have altered their trade preferences, shifting them away from protectionism even in times of economic distress. . . . Intra-industry disputes have weakened pressures for protection." Milner found the French state "in constant negotiation with its industries. . . . Its policy makers did not appear so isolated or independent, and its policy was less coherent and long-term oriented" than other accounts had suggested. "Rising international economic interdependence has been experienced similarly in the United States and France" (pp. 288–89).

Be that as it may, rational-choice analysts are increasingly acknowledging the importance of institutional differences, as Baldwin does by incorporating their features as supplements to a basic public choice model. However much insight the institutionalist perspective may contribute, though, its advocates rarely offer it as sufficient by itself in the trade area. As G. John Ikenberry acknowledges, "in terms of explaining foreign economic policy, the institutional approach can provide only some of the answers. The approach focuses on the structural constraints that confront social and governmental actors in the process of policymaking. Yet while constraints tell us what is not possible, they provide less guidance concerning what is."[21] Destler's account, too, blends in relevant insights of micro- and macroeconomics and makes cognitive arguments as well.[22]

Studies emphasizing this perspective, much like cognitive analyses, typically pay for these broader insights by deliberately sacrificing some precision and generality. Many proceed by induction and avoid quantitative measures and precise general propositions. Baldwin, however, shows ways of making institutional arguments more precise and testable.[23]

[21] Ikenberry, "Conclusion: An Institutional Approach to American Foreign Economic Policy," *International Organization* 42 (Winter 1988), 242.

[22] For other recent examples of synthesis, see David Mares (fn. 17), and Joanne Gowa, "Public Goods and Political Institutions: Trade and Monetary Processes in the United States," *International Organization* 42 (Winter 1988), 15–32.

[23] Other recent research on American government is also developing institutionally richer models. See, e.g., Kenneth A. Shepsle and Barry R. Weingast, "The Institutional Foundations of Committee Power," *American Political Science Review* 81 (March 1987), 85–104, and

THE PERSPECTIVE OF INTERNATIONAL STRUCTURE

A fourth perspective turns trade policy outside in, viewing it as a product of the international structure of which the state is a unit, or as a response to specific interactions between states. David Lake's ambitious study, *Power, Protection, and Free Trade*, defends one such argument: "The central proposition of this book is that the international economic structure and the position of nation-states within it create constraints and opportunities that shape the trade strategies of countries in important and predictable ways" (pp. 29–30). Lake has learned from the debate on economic hegemony, but he considers it too confining because it fails to differentiate among nonhegemonic structures. His interesting refinement proposes a two-dimensional notion of international structure, situating every state relative to others according to economic size and labor productivity.[24] From these properties, Lake attempts to deduce the national trade interests of states in each sector of the grid.

A self-proclaimed follower of Alexander Hamilton, Friedrich List, and Joan Robinson, Lake represents a tradition long in dispute with the liberal one, and his book fires off a challenge that matches in boldness the opposite demarche of Magee and Young. Agreeing that states often violate Adam Smith's principles, he maintains that this is not because of irrationalities due to domestic politics, but precisely because states are rational. Lake's stance rests on several alternative premises: the international system is anarchic; nation-states—rather than individuals or global society—are the most important actors and the appropriate unit of analysis; a state will be concerned about its position relative to other states, as well as about its absolute economic progress; both power and wealth are proper goals; and protection and free trade are both legitimate and effective policy instruments for achieving national interests (p. 2). This realist or statist interpretation concentrates heavily on a single causal factor. Though Lake concedes some importance to internal politics, he deliberately pushes it from the center to the periphery of the analysis. He also raises doubts, from the empirical record, about export dependence, interest group struggle, and political party alternation as

David P. Baron and John A. Ferejohn, "Bargaining in Legislatures," *American Political Science Review* 83 (December 1989), 1,181–1,206.

[24] For other recent reconsiderations of the hegemony debate, see Bruce Russett, "The Mysterious Case of Vanishing Hegemony," *International Organization* 39 (Spring 1985), 207–32; Duncan Snidal, "The Limits of Hegemonic Stability Theory," *International Organization* 39 (Autumn 1985), 579–614; and Joanne Gowa, "Rational Hegemons, Excludable Goods, and Small Groups: An Epitaph for Hegemonic Stability Theory?" *World Politics* 41 (April 1989), 307–24.

explanations of the broad changes in U.S. trade strategy over the decades, while conceding their importance for more specific policy variations.

Lake relies on three hypotheses from international economics to develop a case for protection under certain conditions: optimal tariff theory, the "new mercantilism" focus on leakage of national income through spending on imports, and the benefits of specializing in industries with dynamic increasing returns to scale. Two central conditions, he maintains, can make protection attractive. First, a nation's relative economic size (its share of world trade) affects its ability to benefit from a non-zero optimal tariff and to manipulate the open-economy multiplier. Second, nations with low relative labor productivity are said to have a particular interest in infant-industry or "increasing-returns" protection, unless their domestic markets are small.

These two related dimensions then become axes for defining simple models of different international economic structures. A hegemonic structure, like that prevailing at the turn of the century and in 1950, is characterized by a single large and relatively productive national economy. Lake maintains that around 1912 Britain slipped below a theoretical line between large and medium-sized, and so changed the world structure from a hegemonic one to what he dubs "bilateral opportunism," meaning one with two, and only two, medium-sized, relatively efficient countries. This structure held until the early 1930s, when the British economy also slipped across the average productivity line, leaving the United States alone as both medium-sized and still relatively efficient, and creating a second nonhegemonic structure he calls "unilateral opportunism."

From these systemic classifications the author deduces the changing strategies the United States should have chosen in each period. He uses simple two-by-two game-theoretic models to do so, and in this respect the book provides a sounder game theory application than is common in political science. The theoretical logic furnishes a basis for preference orders prior to the play of a game, thus avoiding the all-too-frequent circularity problem. Most of the rest of the volume then presents an extended empirical defense of this theory, attempting to show that even an unlikely case—the self-sufficient, pluralist, isolationist United States—shifted its external trade strategy rationally in response to changes in the international system between 1887 and 1939.

Lake gives us a better grasp of the sweep of change in broad trade strategy over the decades than is available in much of the literature. He finds theory and evidence related in reasonable, at times even uncanny,

correspondence to each other through much of the period. Until 1912, American global incentives indicated free riding, and industrial tariffs were high. In 1913, an increasingly competitive United States reduced its tariffs dramatically. The subsequent situation is seen as a Prisoner's Dilemma, in which the two dominant players can reach either cooperation or conflict, depending on the degree of international price and exchange-rate stability—an exogenous market variable. Following the war, greater instability caused both countries to increase protection. (Lake assumes greater government protection is the only logical response to instability; he does not consider the rival premise that a declared policy of nonintervention and encouragement of private stabilization would be the most cost-effective route to stability.)

This quest to prove all American actions rational faces a difficult challenge with Smoot-Hawley, regarded almost universally as an unambiguous disaster for broader U.S. interests. Lake shores up the theory with a clever amendment, which holds that if players were to foresee a cooperative game coming to an end, preemptive defection would pay. If everyone will soon be beggaring like mad, the one that moves first will gain briefly, until the others have retaliated. Britain slipped across the average productivity line at the end of the 1920s, and Lake maintains that America must have seen this as the end of one structure and the coming of another with fewer constraints. The British were already moving toward protection during the twenties, and Lake contends that this undermined the U.S. interest in avoiding retaliation; new barriers were coming anyhow. Inconvenient for this interpretation, of course, is the fact that the decisive British turn to imperial preference did not occur until two years after Smoot-Hawley. Lake does concede that U.S. domestic politics got out of hand, despite some remaining systemic incentives for restraint.

Unlike most political scientists, Lake ventures explicit predictions for the future. He rejects the 1930s analogy for today, because we find ourselves in a different nonhegemonic structure.

> If the theory developed above is correct, the international economy will remain relatively open and liberal despite the decline of American hegemony. The international economy will resemble that which existed between 1912 and 1930, but we are unlikely to see a repetition of the economic conflicts of the 1930s. (p. 229)

He feels relatively optimistic because the present system has not one but several productive economies of middle size, all of which have interests in open trade as well as some influence. At the same time he concedes

that the larger number raises the problem of free riding, and that enforcement by retaliation is risky. Unfortunately, Lake stops before he has elaborated a convincing contemporary constraint on the buck-passing problem that has often been blamed for the 1930s trade war. We might find such constraints, however, by introducing the greatly thickened web of interdependence at the industry level, expressed via domestic politics, and causal beliefs based on the depression experience, emphasized in the first and second perspectives.

Lake's stimulating book casts a new light on the events of those years and should be debated seriously. Like some of the narrower liberal treatments, however, this relatively narrow antiliberal theory is vulnerable to several objections, as the author is aware. First, many readers will have doubts about the assumption that the benefits of protection to the U.S. economy exceeded its costs, and they will search in vain for data to substantiate it. In fairness, though, it must be added that Lake's method is analogous to a common position taken by liberal theorists: question not the accuracy of my assumptions; look instead at how productive their implications are.

Second, heavy interpretive weight is placed on particular operationalizations that are debatable. As Milner points out, for instance, an equally good case can be made that U.S. power was already hegemonic by 1930. The American share of world trade had almost reached the level it was to attain in 1950, and other data—such as shares of world production—point in the same direction. In any case, it is difficult to agree that the future of world trade hinges on whether the U.S. share is 14 percent or 17 percent. Third, the theory itself faces empirical anomalies and is not broad enough to carry the entire explanatory burden, as Lake acknowledges (p. 224).

Generally, systemic structural theory has shown that it has much to contribute to the understanding of international economic relations. This version in particular may also fit even better in those many cases that are not as unlikely as that of the United States. It would stand to reason that smaller, more open political economies will shift their strategies to a greater extent in response to external incentives. Considerable careful research does confirm that strong and weak powers behave differently in trade relations, and that weaker ones tend to make concessions under external pressure, though even they have room to maneuver in response to their varying doctrines and domestic political conditions.[25]

[25] See Tony Smith, "Changing Configurations of Power in North-South Relations since 1945," *International Organization* 31 (Winter 1977), 1–28; Stephen D. Krasner, "The Tokyo Round: Particularistic Interests and Prospects for Stability in the Global Trading Regime,"

STRENGTHENING THE SYNTHESIS

To sum up, recent scholarship on U.S. international trade policy is a tapestry constructed of several distinct analytical traditions and has not attained theoretical unity. From the evidence, each of four analytical perspectives—emphasizing market conditions, policy beliefs, political institutions, and global structures—has made valuable contributions. Yet each by itself also proves inadequate as a sole unifying vehicle, so that there is clearly much to be gained by moving further toward a more comprehensive and integrated view of the issue.

Such a synthesis is in fact developing. Though scholars will always disagree on some dimensions, note that not all the books under review fall neatly into two traditional boxes, those by economists versus those by political scientists. While each author does call one of those disciplines home, most have also begun to study scholarship produced by the other, and have deployed insights gained from the exchange, albeit selectively. For example, economists Baldwin and Bhagwati cite and contribute to political science literature; political scientist Milner defends a conclusion having as much or more in common with the market approach as with power and institutional arguments. Increasingly, analysts are characterizing their studies explicitly as contributions to a joint political/economic science, and expect to be judged by more than one conventional discipline. Most of the recent books under review, in addition to other studies mentioned in passing, blend insights from several perspectives. To put the point another way, the classification of thought in political economy as either liberal, mercantilist, or Marxist is becoming less and less useful. The literature under review, at any rate, is refusing to stay within those categories.

What are the key features of this emerging integrated theory of trade policy formation? It clearly encompasses individuals, groups, and states as actors—rejecting the argument of parsimony for excluding any of them. Discerning the effects of each on the others is therefore central to the complex analytical challenge. How they combine is becoming apparent in particular episodes and arenas, but is still not clear in general. The consensus theory assumes that decisions are made rationally, in the sense of process rather than substance. Preferences, however, are not taken as given; the theory recognizes that preferences are partly shaped according to ideologies, theories, and political institutions and processes that vary

International Studies Quarterly (December 1979), 491–531; Peter J. Katzenstein, *Small States in World Markets* (Ithaca, NY: Cornell University Press, 1985); and John S. Odell, "The Outcomes of International Trade Conflicts," *International Studies Quarterly* 29 (September 1985), 263–86.

across countries and change through time. This essay also highlights several more specific propositions on which consensus is developing.

The emerging theory naturally excludes some orientations. At the level of epistemology, this community is neopositivist without apology; no poststructuralist or postmodernist dissent is evident here.[26] Marxist thinking is also absent.

At the level of broad methodology, we are seeing a welcome, gradual movement away from efforts to defend single inadequate theories and vanquish all others, and toward joint learning and integration. To be sure, there will always be an honored place for partial analysis, if developed as a means to a larger end. Furthermore, efforts to defend provocative theses or antitheses may have heuristic value even if they are eventually abandoned. Freshness and controversy will always be valued, and criticism from outside is certainly needed to prevent any consensus from lapsing into stagnation.

It is less often noted, however, that provocation and novelty have their costs as well. Being wrong grandly also consumes resources and sometimes generates wasteful debating cycles, which can become more ideological than scientific. Political economists, for example, have known debates between neoclassicists, statists, and Marxists that rehearse familiar arguments for the superiority of their respective approaches without moving understanding forward either theoretically or empirically. Partisans will no doubt continue to attempt to vanquish their critics, and some will be tempted to do so by insisting on a partial or even an extreme alternative. But this field as a whole will benefit more over the long term if new theories are developed in the spirit of aggressively seeking to incorporate valuable insights generated by alternative traditions rather than by dismissing them without investigation.

In the next stage of research on trade policies, two major substantive projects would be especially fruitful. One would broaden this literature by reducing its concentration on the Unites States. We need more comparable studies of the rest of the world that are sensitive to the institutional differences among countries. The second would investigate more fully the interaction between national trade policies and international relations.

We have some English-language studies of other countries, but not many comparative analyses of policy formation. Britain's repeal of the Corn Laws and other decisions, such as entering the Common Market,

[26] Cognitive interpretations might seem to introduce some interpretive or phenomenological elements, but in many cases they are offered as primitive social-scientific hypotheses to be sharpened and tested in principle in the same manner as other hypotheses.

have been chronicled.[27] There are historical and more recent studies o German[28] and Canadian[29] policy, and works on Japan have been mentioned.[30] A few cross-national comparative studies have also been completed,[31] but those using the same research design systematically to explain policies of two or more countries are still rare. Particularly notable is the scarcity of work on the EC as a global player. The European Community, one of the world's three trading giants, operates as a unit in international trade negotiations, but a reader who wants to understand why it makes external trade policy as it does, and responds to demands or offers as it does, will have difficulty locating comparable recent studies, at least in English.[32]

Trade policy is only one strand of economic policy, and analyzing it as a component of a more complex package is also illuminating. We do have cross-national comparative studies of industrial, agricultural, and macroeconomic policies that encompass import and export barriers and subsidies to varying degrees. Because they ask broader questions, however, their conclusions cannot easily be compared with those of the trade

[27] See Charles Kindleberger, "The Rise of Free Trade in Western Europe, 1820–1875," *Journal of Economic History* 35 (1975), 20–55; Norman McCord, *The Anti-Corn Law League, 1838–1846* (London: Allen & Unwin, 1958); J. H. Richardson, *British Economic Foreign Policy* (New York: Macmillan, 1936); Robert J. Lieber, *British Politics and European Unity* (Berkeley: University of California Press, 1970).

[28] Percy Ashley, *Modern Tariff History. Germany-US-France*, 3d ed. (New York: Dutton, 1926); Wilhelm Röpke, *German Commercial Policy* (London: Longmans, Green, 1934); H. Glissman and F. Weiss, "On the Political Economy of Protection in Germany," Staff Working Paper No. 427 (Washington, DC: The World Bank, October 1980). Companion World Bank papers cover several other countries.

[29] Richard E. Caves, "Economic Models of Political Choice: Canada's Tariff Structure," *Canadian Journal of Economics* 9 (May 1976), 278–300; G. K. Helleiner, "The Political Economy of Canada's Tariff Structure: An Alternative Model," *Canadian Journal of Economics* 10 (May 1977), 318–26.

[30] See fn. 3; also H. Richard Friman, "Rocks, Hard Places, and the New Protectionism: Textile Trade Policy Choices in the United States and Japan," *International Organization* 42 (Autumn 1988), 689–724.

[31] E.g., Peter Katzenstein, ed., *Between Power and Plenty* (Madison: University of Wisconsin Press, 1978); John A. C. Conybeare, "Tariff Protection in Developed and Developing Countries," *International Organization* 37 (Summer 1983), 441–68; Gallarotti (fn. 9).

[32] A sampling of recent works that are related in some way could include Chihiro Hosoya, "Relations between the European Communities and Japan," *Journal of Common Market Studies* 18 (December 1979), 159–74; E.L.M. Volker, ed., *Protectionism and the European Community* (Deventer, The Netherlands: Kluwer Law & Taxation Publishers, 1983); Helen Wallace, William Wallace, and Carole Webb, eds., *Policy Making in the European Community*, 2d ed. (New York: John Wiley, 1983); Miles Kahler, "European Protectionism in Theory and Practice," *World Politics* 37 (July 1985), 475–502; Meinhard Hilf, Francis G. Jacobs, and Ernst-Ulrich Petersmann, eds., *The European Community and the GATT* (Deventer, The Netherlands: Kluwer Law & Taxation Publishers, 1986); Winham (fn. 4), passim; Robert E. Baldwin, Carl B. Hamilton, and André Sapir, eds., *Issues in US-EC Trade Relations* (Chicago: University of Chicago Press, 1988); Georg Koopmann and Hans-Eckart Scharrer, "EC Trade Policy beyond 1992," *Intereconomics* 23 (September/October 1989), 207–15; and works cited earlier.

policy literature. Still, we might learn more about the United States itself *if we viewed it more often through these broader lenses. American scholars most accustomed to their relatively liberal state find it natural to focus on "trade policy" rather than "industrial policy,"* even though the United States actually intervenes more in its economy than official U.S. doctrine reveals. Enlarging the focus more often, to document and explain U.S. industrial and agricultural policies in their entirety and in comparative perspective, would throw new light on hypotheses designed to explain only trade policies.

Another growing literature of the 1980s and 1990s investigates different developing countries: their reasons for following diverging economic strategies, their different responses to debt and balance-of-payments problems, and their stabilization experiences and liberalization, including trade liberalization. Here, too, the more comprehensive role of the state has encouraged a research question broader than trade policy alone. It would be interesting to see more comprehensive comparative studies of these patterns of accomplishing (or not accomplishing) adjustment to international payments problems, with the United States as one of the cases.

The second project would establish better links between trade policy research and the rest of international relations. Many works in this field have abstracted from overseas influences, concentrating on the domestic sources of policy. Even if this simplification was understandable in the case of the Unites States in the past—a premise questioned in several of the books under review and in others—the trend toward opening the U.S. economy makes it increasingly dubious even for America. To what extent can the trade policies of the United States and other nations be understood as a function not only of the international power structure or increasing global economic integration, but also of particular trade negotiations, threats, or inducements from other states?

In addition, while many commentators affirm that trade and economic policies are inextricably intertwined with security conditions, scholarship has not rigorously documented such trade-security links. It is true that substantial subliteratures have investigated the relationship between military spending and economic growth, long historical cycles of war and production, East-West trade, and the use of economic sanctions for political purposes.[13] Robert Stern's book includes an interpretive

[13] For windows into those topics, see, respectively Adne Cappelan, Nils Peter Gleditsch, and Olav Bjerkholt, "Military Spending and Economic Growth in the OECD Countries," *Journal of Peace Research* 21 (November 1984); 361–73; Riccardo Faini, Patricia Annez, and Lance Taylor, "Defense Spending, Economic Structure and Growth: Evidence among Countries and over Time," *Economic Development and Cultural Change* 32 (April 1984); 487–

survey of "Trade Policy as Foreign Policy" from 1765 to 1985 by Richard Cooper (pp. 291–322) and a critique of "The National Defense Argument for Government Intervention in Foreign Trade" by T. R. Srinivasan (pp. 337–63). But are trade policies toward allies systematically more favorable than those toward non-allies, or are variations even greater among military allies than between them and non-allies? Are some sectors more insulated from security linkages and others more sensitive? Are security linkages more powerful in periods of military crisis than they are otherwise? We have little systematic comparative research to answer questions such as these.[34]

These two projects will need contributions from each of the analytical traditions. For instance, the market-based theory could test the sensitivity of its micro and macro findings to changes in the national institutional environment by investigating the same questions in Japan, Brazil, or southern Africa. Research on how the theory relates to international trade bargaining—meaning coercive, retaliatory moves as well as accommodative compromises—would be equally interesting. We know that international market conditions can change the structure of organized interests that are ready to bring pressure to bear domestically, but we also know that governments (in Japan, for example) may be facing external negotiations at the same time. Just how these changing market forces, internal bargaining, and international negotiations all affect each other is not clearly established anywhere.

Scholars could advance the cognitive perspective by reflecting on the fate of the liberal trade doctrine during the 1970s and 1980s in the United States, where it appeared to lose some ground to notions of fair trade and industrial policy just as it seemed to be gaining adherents south of the border and in the East. How much damage can interests sustain before ideas give way, to be displaced by alternative ideas? More generally, under what conditions, cross-nationally, do publics abandon one doctrine for another, and which alternative beliefs will be favored by what conditions? Can we gather better evidence on leaders' and publics' changing

98; Joshua Goldstein, *Long Cycles: Prosperity and War in the Modern Age* (New Haven: Yale University Press, 1988); Franklyn D. Holzman and Robert Legvold, "The Economics and Politics of East-West Relations," in C. Fred Bergsten and Laurence B. Krause, eds., *World Politics and International Economics* (Washington, DC: The Brookings Institution, 1975); David Baldwin, *Economic Statecraft* (Princeton: Princeton University Press, 1985); and Hufbauer et al. (fn. 2).

[34] K. J. Holsti, in "Politics in Command: Foreign Trade as National Security Policy," *International Organization* 40 (Summer 1986), 643–72, suggests reasons for such variations. James Alt and Barry Eichengreen formalize such linkages as "parallel games," and explore their implications in "Parallel and Overlapping Games: Theory and an Application to the European Gas Trade," in Odell and Willett (fn. 16).

attitudes toward security and economic goals in a number of countries, and use it to reach better assessments of the effects of these attitudes in economic policy formation?

The proposition that different cultural histories and national political institutions shape individual and industry preferences differently despite similar market conditions could be supported with much more systematic empirical research, either quantitive or using small-n case contrasts. Carefully designed studies of historical institutional changes as well as cross-national comparisons could investigate how industry's political preferences shift and political coalitions form and reform around economic policies. Over the longer term (President Gorbachev might ask), what market conditions are sufficient to change national institutions fundamentally, in what respects?

The international level of analysis has generated much work not considered here, and some of it could usefully be related to these questions. Bhagwati, among others, makes the argument that the existence of an international regime, the GATT, has influenced trade policies for the better since 1947 (p. 41), but this remains largely an assertion. We know that the GATT has coincided with tariff reductions as well as with the spread of new forms of protection, and we suspect there are causal relations here, but it is difficult to isolate the GATT's effects from other factors that would have operated in any case. Generally the relationship between regimes and state policies has proved to be a highly elusive one to document empirically;[15] like cognitive arguments, it requires more careful work before it can stand up to rigorous scrutiny. Similarly, the relationship between international negotiation and state policies, outside of regimes as well as within them, may prove to be a fruitful source of insight.

One might reflect that it is fine to endorse openmindedness and empirical research, but what about parsimony? Could this call for synthesis become a recipe for an endless proliferation of theories, ultimately undermining the value of theory itself as a simplifying device? Not necessarily. Simplicity is certainly valuable; most analysts would welcome a powerful, spare model that also proved equally valid empirically. The problem is that none of the available alternatives has come close to achieving this ideal.

We need to reduce complexity and theoretical proliferation, but by encompassing rather than ignoring significant dimensions and results. One indirect approach to this perennial dilemma of parsimony versus rich empirical validity would be to develop *conditioning hypotheses*, which

[15] See Stephan Haggard and Beth A. Simmons, "Theories of International Regimes," *International Organization* 41 (Summer 1987), 491–517.

specify the conditions under which other hypotheses are most likely to be useful or valid. Say, for example, that we are interested in the level of import protection—call it variable *X*—and we have several primary variables and hypotheses—call them *A*, *B*, and so on—for explaining protection. We also have inconsistent empirical results for some of these explanatory ideas. This apparent confusion can be reduced if we can identify other variables and hypotheses—*M* and *N*—specifying general conditions under which the effects of variable *A* on *X* are amplified or suppressed. With a conditioning hypothesis in hand, we can explain why the results are positive in some studies and not in others.

To illustrate briefly, suppose *A* is the degree to which the government sees a security interest in trade liberalization, and *B* is the distribution of domestic interest group pressures for and against protection. Hypothesis *M* might identify the centralization of national tariff-setting institutions as conditioning the effects of *B* and perhaps *A* as well. The 1934 change in U.S. institutions moved decisions on specific products to the executive, which made it more difficult for interest groups to make pressures effective, and probably also amplified the effects of executive security goals on trade policy.[16] This illustration also suggests that some conditioning hypotheses may be found among the present menu of "primary" propositions. If so, the simplifying power of such research will be greater still.

Take, for another example, the proposition that a decline in an industry's market share and the resulting group lobbying will make trade policy more protective. This proposition should be less useful when the policy at stake is more like a collective good, and more useful when the good is excludable.[17] For instance, U.S. policy concerning the nature of GATT institutions is typically more like a collective good for Americans, in contrast to policy regarding imports of particular items such as copper or motorcycles.

This strategy of seeking conditioning hypotheses can be pursued by means of both formal deduction and more informal and inductive methods. Modelers and case-study analysts alike can work toward greater power and clearer structure in the emerging theory.[18] If some ordering ideas were shown—not merely asserted—to have greater simplifying

[16] Lake (p. 225) notes historical cross-national contrasts supporting the importance of state centralization.

[17] See Joanne Gowa, "Public Goods and Political Institutions: Trade and Monetary Policy Processes in the United States," *International Organization* 42 (Winter 1988), 15–32. On the other hand, changes in the structure of trade interests over the long term might also reshape the institutions that permit this result.

[18] This suggestion is elaborated further in Patricia Dillon, John S. Odell, and Thomas D. Willett, "Future Directions in the Analysis of Trade Policies," in Odell and Willett (fn. 16), 273–83.

power than others, then we could look toward a hierarchy of hypotheses. Such a hierarchy (which is a useful but probably unattainable ideal) would consist of a set of master or "higher" variables that "regulate" the effects of others attached to them, arranged symbolically in an inverted V pattern. In any case, a structure does seem to be crystallizing in this literature. Concentration on this methodological strategy would accelerate this development, and help us enjoy the benefits of a more comprehensive theory while reducing the unavoidable tension between parsimony and rich understanding.

The extent to which we will continue to enjoy benefits (and costs) of international trade itself for another decade will of course depend on many forces beyond scholars' ideas, though such studies might have some effect in subsequent history. In any case, we can be sure that today's lively trade politics will supply much new material for a fresh assessment a decade hence.

COUNTERFACTUALS AND HYPOTHESIS TESTING IN POLITICAL SCIENCE

By JAMES D. FEARON*

Without the prior democratic modernization of England, the reactionary methods adopted in Germany and Japan would scarcely have been possible. Without both the capitalist and reactionary experiences, the communist method would have been something entirely different, if it had come into existence at all.

—*Barrington Moore*

Nuclear weapons did not cause the condition of bipolarity. . . . Had the atom never been split, [the U.S. and the Soviet Union] would far surpass the others in military strength.

—*Kenneth Waltz*

THE epigraphs[1] provide examples of counterfactual conditionals, or propositions that take the generic form "If it had been the case that C (or not C), it would have been the case that E (or not E)." Counterfactuals make claims about events that did not actually occur. It is argued in this paper that such propositions play a necessary and fundamental, if often implicit and underdeveloped, role in the efforts of political scientists to assess their hypotheses about the causes of the phenomena they study. Particularly in small-N research designs, scholars in comparative politics and international relations routinely evaluate causal hypotheses by discussing or simply referring to *counterfactual cases* in which a hypothesized causal factor is supposed to have been absent. Though this procedure is quite common, its methodological status and its viability are unclear and are worth exploring. How does the strategy of counterfactual argument relate, if at all, to methods of hypothesis testing based on the comparison of actual cases, such as regression analysis or J. S. Mill's Method of Difference? Are counterfactual "thought experiments" a vi-

* Prepared for delivery at the Annual Meetings of the American Political Science Association, San Francisco, August 30–September 2, 1990. For helpful comments and advice I wish to thank Chris Achen, Robert M. Adams, David Collier, Russ Faeges, Don Green, Marcus Kurtz, Jim Mahon, Merrill Shanks, Laura Stoker, Arun Swamy, and an anonymous referee. I am particularly indebted to David Collier for his encouragement and close reading of several drafts. None mentioned (or unmentioned) are responsible for my mistakes.

[1] Moore, *Social Origins of Dictatorship and Democracy* (Boston: Beacon Press, 1966), 414; Waltz, *Theory of International Politics* (New York: Random House, 1979), 180.

World Politics 43 (January 1991), 169–95

able means of assessing hypotheses about national and international outcomes, or are they methodologically invalid in principle?

The article has three principal aims. First, I seek to support the claim made above, that counterfactual propositions and arguments play a central role in the efforts of political scientists to assess their causal hypotheses. Support is drawn from methodological argument and from examples showing the counterfactual strategy at work in research in the areas of comparative politics and international relations. I discuss examples from the literature on the causes of World War I, the nonoccurrence of World War III, social revolutions, the breakdown of democratic regimes in Latin America, and the origins of fascist and corporatist regimes in interwar Europe.

Second, the paper examines how the strategy of counterfactual argument is related to but also differs from methods of hypothesis testing based on the comparison of actual cases. The two approaches are found to be closely related: analysts with few cases and many variables are compelled to resort to counterfactual argument by a *statistical* principle; and counterfactuals also appear to play a key role in the assumptions that justify large-N regression analysis, when the data employed is quasi-, or nonexperimental.[3] The difference between regression and the counterfactual strategy is not that one relies on counterfactuals while the other does not. Rather, the strategies differ in the way that each employs counterfactuals and in the way that each evaluates support for a causal hypothesis.

Finally, the paper addresses the question posed above: Is counterfactual argument a viable means of assessing causal hypotheses in nonexperimental research settings? I give no firm answer here but instead introduce some of the problems and issues involved. My purpose is neither to advocate the use of counterfactual argument in preference to comparisons with other actual cases nor to suggest that the counterfactual strategy is fundamentally invalid or has no value in principle. Since political scientists often use counterfactual argument when assessing or justifying causal hypotheses, particularly in small-N research settings, it seems important simply to understand what the strategy entails. If the paper carries a methodological prescription, it is that researchers who use counterfactual argument to support causal hypotheses should be methodologically aware of what they are doing and should make their counterfactual arguments as explicit and defensible as they can.

These general aims are pursued in three sections. The first distin-

[3] Data not generated by random assignment to control and treatment groups is referred to as quasi-, or nonexperimental.

guishes between two strategies of hypothesis testing—the comparison of actual cases and counterfactual argument—and examines key methodological similarities and differences between them. The second section shows how the counterfactual strategy appears in practice by considering examples from work in international relations and comparative politics. The examples make clear that counterfactuals matter both when the researcher is focusing on one actual case (for example, the outbreak of World War I or the Brazilian military takeover in 1964) and when the researcher considers several actual cases (for example, social revolutions or interwar European regime types). The third section returns to some theoretical issues concerning the link between causal arguments and counterfactual propositions, issues that bear on the question of whether the counterfactual strategy is a viable one. Two logical problems related to the use of counterfactuals are discussed briefly: (1) Is any event C that appears to satisfy "if C had not occurred, E would not have occurred" to be called a "cause" or E? And (2) Are some counterfactual comparisons more "legitimate" or appropriate than others?

Counterfactuals, Actual Case Comparisons, and the Logic of Inference

Suppose it is hypothesized that C was a cause of event E. I would argue that when experimental control and replication are not possible, analysts have available a choice between two and only two strategies for "empirically" assessing this hypothesis. Either they can imagine that C had been absent and ask whether E would have (or might have) occurred *in that counterfactual case*; or they can search for other *actual* cases that resemble the case in question in significant respects,[3] except that in some of these cases C is absent (or had a different value). In the latter procedure, the analyst then checks the association between the occurrence of C and E in the set of actual cases.[4] If successful (from the analyst's point of view), both strategies would tend to support the hypothesis that the proposed cause in fact produces (or produced) the effect.

As an illustration, consider the hypothesis that international structural rather than domestic political factors have been the principal causes of

[3] The sense of "significant respects" is discussed below.

[4] These summary statements of the two strategies are not complete. Qualifications and elaborations for each are discussed in the rest of the paper, with more attention paid to the counterfactual case strategy. The potential difficulties with the method of comparing actual cases, which is formally known as regression analysis though informally practiced in such works as Theda Skocpol, *States and Social Revolutions* (Cambridge: Cambridge University Press, 1979), are extensively discussed in the econometrics and statistics literatures.

major aspects of Soviet foreign policy. The analyst applying the counterfactual strategy would evaluate the hypothesis by examining arguments that *any* regime in Russia, Soviet or not, *would have* made essentially the same foreign policy choices. The analyst using the actual case strategy would search for cases of states in both similar and dissimilar structural positions as Soviet Russia and then would check the sample for a relationship between structural position and foreign policies.[5]

Both methodological strategies aim to solve *the same statistical problem*. Our analyst begins with one case and at least one explanatory variable, which means negative degrees of freedom.[6] Legitimate causal imputations cannot be made on the basis of negative degrees of freedom, so the analyst wishing to assess a causal hypothesis or to assess the relative weights of different causes has no choice but to add or create more cases: either a counterfactual case (or cases) that never actually existed or actual cases.

Put otherwise, the analyst, in explaining why some particular event E occurred, cannot help but explain why E occurred *rather than* some other possible outcome or outcomes. These other possible outcomes define the range of variation that the analyst accounts for, and this range is treated differently in different research traditions. For example, much historical analysis leaves implicit the other things that might have been had the historian's favored causes varied. In more methodologically self-aware small-N work, analysts tend to be more explicit about what might have happened.[7] Finally, in the actual case strategy, analysts take their cues about what might have happened from other actual cases. Thus, an elections specialist may explain why a respondent voted Republican rather than Democratic (as did other actual respondents); students of international conflict may explain why deterrence failed in one actual case but not in other cases; experts in comparative politics may explain why interwar Germany became a fascist dictatorship rather than a liberal de-

[5] "Structural position" here would entail the number of great powers and the basic geopolitical circumstances of the Soviet Union. Waltz (fn. 1); and idem, "Another Gap?" in Robert Osgood et al., *Containment, Soviet Behavior, and Grand Strategy*, Policy Papers in International Affairs No. 16 (Berkeley: Institute of International Studies, University of California, 1981). On structural versus domestic political or ideological explanations of Soviet foreign policy, see also Barry R. Posen, "Competing Images of the Soviet Union," *World Politics* 39 (July 1987), 579–97.

[6] Degrees of freedom are the number of cases minus the number of explanatory variables minus one.

[7] For example, Barrington Moore, *The Social Bases of Obedience and Revolt* (London: Macmillan, 1978); Stephen Van Evera, "The Cult of the Offensive and the Origins of the First World War," *International Security* 9 (Summer 1984), 58–107. Of course, historians can be quite careful about their counterfactual arguments. For examples, see McGeorge Bundy, *Danger and Survival* (New York: Random House, 1988); and George Kennan, *Russia and the West under Lenin and Stalin* (Boston: Little Brown, 1960), 29–32.

mocracy like England or (taking a larger range of other actual cases) a social democracy like Sweden or a traditional dictatorship like Austria in the Dollfuss-Schuschnigg period.[8]

Not only do counterfactual and actual case strategies both attempt to solve the same statistical problem, but both also run important methodological risks. Less obvious is that in each of the strategies, the principal risks are closely connected to the role played by counterfactuals.

The main risk in the first strategy is obvious and serious—how can we know what would have happened with any degree of confidence? Historians, when confronted with the suggestion that the validity of their causal inferences *necessarily* depends on counterfactual argument, have often dismissed out of hand or ignored the idea in favor of the view that their job is to deal with *reality*.[9] Political scientists and sociologists, too, with the exception of a neglected methodological piece by Max Weber and some recent work by Jon Elster, have also tended to avoid explicit discussion or open embrace of the counterfactual strategy, probably because it is felt that an empirical political science must deal only with actual cases. This belief would seem to be reflected in the title of a recent book of essays by political scientists working with counterfactual premises: *What If?: Essays in Social Science Fiction*. The play on "science fiction" is no accident here.[10]

The risks of the second methodological strategy—that of increasing the degrees of freedom by considering other actual cases—are also well known. Whereas in the counterfactual approach one tries to imagine another (not actual) case in which the presumed causal agent is absent but everything else that is relevant is identical, in the second strategy the analyst adding actual cases may not know if the additional cases are appropriately identical. If there are other causes of the phenomenon in

[8] On deterrence, see Paul Huth, *Extended Deterrence and the Prevention of War* (New Haven: Yale University Press, 1988); on fascism versus liberalism in Germany, see Moore (fn. 1); on fascism versus liberalism, corporatism, or traditional dictatorship, see Gregory M. Luebbert, "Social Foundations of Political Order in Interwar Europe," *World Politics* 39 (July 1987), 449–78.

[9] For example, according to A. J. P. Taylor, "a historian should never deal in speculations about what did not happen"; Taylor, *The Struggle for Mastery in Europe, 1848–1918* (London: Oxford University Press, 1954), 513. Or, in M. M. Postan's words, "The might-have-beens of history are not a profitable subject of discussion"; quoted in J. D. Gould, "Hypothetical History," *Economic History Review*, 2d ser., 22 (August 1969), 195–207. See also David Hackett Fischer, *Historians' Fallacies* (New York: Harper Colophon Books, 1970), 15–21; and examples given in Peter McClelland, *Causal Explanation and Model-Building in History, Economics, and the New Economic History* (Ithaca, N.Y.: Cornell University Press, 1975).

[10] Weber, "Objective Possibility and Adequate Causation in Historical Explanation," in *The Methodology of the Social Sciences* (New York: Free Press, 1949); Elster, *Logic and Society: Contradictions and Possible Worlds* (New York: Wiley, 1978); citations below (fn. 56); and Nelson Polsby, ed., *What If?: Essays in Social Science Fiction* (Lexington, Mass.: Lewis Publishing, 1982).

question that are not considered explicitly in the analysis, and if any of these are in fact systematically related to the causes explicitly considered, then effects of the other causes will be wrongly attributed to those of the causes that are being evaluated. Simply put, estimates of the effects of the proposed causes will be biased. In statistics this is the familiar problem of whether any independent variables are correlated with the contents of the error term (which contains the effect of all unspecified, unmeasured "other causes"). Such correlation may occur due to failure to include relevant independent variables, errors in measuring the independent variables, or unrecognized reciprocal causation. In the comparative politics literature it is often posed as the question of whether a researcher's several cases are comparable, or if the ceteris paribus assumption is adequately satisfied.[11] Because of the severity of this risk, some analysts tend to be skeptical of large-N or comparative historical work; they prefer case studies in which the risks of (an often implicit) counterfactual strategy may seem intuitively less serious.

Less well understood is the link between this central risk run by the actual case strategy and counterfactuals. While the paper focuses primarily on the role of counterfactuals in small-N research, a few words on their role in quasi-experimental regression analysis are useful as a prelude to making clearer exactly how the two strategies differ.

To support a causal interpretation of estimated regression coefficients, the large-N analyst using nonexperimental data needs to make a number of theory-driven assumptions.[12] As noted, chief among these is the as-

[11] The notion of comparability plays a major role in the methodological and applied writings of specialists in comparative politics. My impression is that nonetheless the notion remains a deeply vague one. It seems to include, at various times, the idea that the other causes should be uncorrelated with the independent variables ($E(X'e) = 0$); that everything else should have as little influence as possible ($E(e'e)$ should be close to zero); that measures will not be as valid or reliable across countries and cultures; and other meanings. (Throughout, $E(\cdot)$ is the expectations operator; X is an $n \times k$ matrix of n observations on k independent variables; e is an $n \times 1$ vector of error terms.)

Posing the main risk for analysis across sets of actual cases in terms of the regression validity of the ceteris paribus assumption also bears qualification. For regression estimates to be unbiased, we do not need the other things to be literally equal, though it is true that the more equal they are, the greater the precision of our estimated effects. For unbiased estimates of causal effects we need only require that the other things not be systematically related to the prospective causes and the dependent variable that we are evaluating. This point appears not to have been fully clear to Mill (working before statistics was well developed), who sometimes writes in his *System of Logic* (London: John W. Parker, 1851) as though everything else has to be literally identical in order for the Method of Difference to work. The same confusion seems to carry over today in the work of some specialists in comparative politics who take Mill as a principal methodological guide (e.g., Theda Skocpol and Margaret Somers, "The Uses of Comparative History in Macrosocial Inquiry," *Comparative Studies in Society and History* 22 [April 1980], 174–97). That said, I should also note that those who conduct large-N research often do refer to the assumption as the "ceteris paribus assumption" simply for convenience, and I will follow this usage here.

[12] This is true as well of actual experiments in which cases are assigned at random to

umption that explanatory variables and the errors (the other causes) are uncorrelated. Formally, the argument that estimated coefficients are unbiased depends on the assumption that $E(X'e) = 0$. It is easy to show that this assumption is credible *if and only if* a counterfactual proposition is credible; namely, the proposition

> *(P1) If the cases in the sample had assumed different values on the independent variables, the contents of the error term would not have differed systematically.*

If P1 *is false, then* $E(X'e)$ *does not equal zero. If* $E(X'e)$ *does not equal* zero, P1 cannot be true (Q.E.D.).

This argument says that assuming that $E(X'e) = 0$ in a quasi experinent is *equivalent* to assuming the truth of a counterfactual proposition ibout what would have happened if we could have altered a variable's 'alue for any case in the sample. Although one may not think about the eteris paribus assumption in terms of a counterfactual proposition, a ounterfactual proposition is necessarily involved nonetheless. In actual xperiments random assignment guarantees the truth of P1 (within the imits of sampling variance). In quasi experiments, a causal interpretaion of estimated coefficients requires belief in the credibility of the counerfactual P1 for justification. If we believe the results of a regression nalysis, we must be willing to believe that, say, if Joe Respondent had ieen a Republican as opposed to a Democrat, he would have been oughly "so much" more likely to have voted for Reagan in 1984; or that f Israel had not moved troops quickly to its northeastern border in Sepember 1970, then Syria would have been much less likely to have been eterred from advancing on the Jordanian capital.[13]

If both strategies of confirmation are means of solving the same statisical problem, and if both depend in some measure on counterfactuals, hen how do they differ? The answer is that each strategy provides its empirical" confirmation for a causal hypothesis in a different way.

In the actual case strategy support for a hypothesized causal connecion comes principally in the form of a frequency or magnitude of assoiation across actual cases. Of course, theory-driven assumptions—mong them a counterfactual one—are needed to support or justify any egression result. But the result *one looks for* in regression analysis is an

---

'eatment and control groups. See Leland Neuberg, *Conceptual Anomalies in Economics and 'atistics: Lessons from the Social Experiment* (Cambridge: Cambridge University Press, 1988). .mong other things, Neuberg shows that a counterfactual assumption is needed to justify .timates of sampling variance in actual experiments. I suggest below, however, that coun'rfactuals play a key role in quasi-experimental hypothesis testing that they do *not* play in 'tual experiments.

[13] The later example comes from Huth (fn. 8), 97, who often uses counterfactual argument oout particular cases to make more plausible the results of his regression analysis.

estimated coefficient significantly different from the null hypothesis, and this difference derives from a frequency of association in the sample.

In the counterfactual case strategy, by contrast, frequencies of association cannot be meaningfully assessed. They are arguably irrelevant in any event, since the researcher is attempting to perform the perfect experiment, in which everything but the test factor is equal. Instead, support for a causal hypothesis in the counterfactual strategy comes from *arguments* about what would have happened. These arguments are made credible (1) by invoking general principles, theories, laws, or regularities distinct from the hypothesis being tested; and (2) by drawing on knowledge of historical facts relevant to a counterfactual scenario.[14]

An example will help make this point concrete. It has been proposed that a "cult of the offensive"—the widespread conviction held by European civilian and military leaders that there were enormous strategic advantages to striking first—was an important cause of World War I.[15] According to the analysis above, there are two means of empirically checking this hypothesis. Following the actual case strategy, we could assemble a set of international disputes, some of which escalated to war and some of which did not. We could then construct a measure of military and civilian beliefs about the advantages of a first strike, presumably from military writings and from statements of politicians and generals about their expectations for war. Finally, after thinking hard about what other independent variables required statistical control, we could test for the strength of association between commitment to offensive doctrines and escalation. To assess the contribution of this cause to the likelihood of World War I in particular, we would check the value of the several independent variables for this case, comparing their various contributions with that of belief in first strike advantages.[16]

Alternatively, we might employ the counterfactual case strategy,

[14] I am relying here on what David Lewis calls "metalinguistic" theories of counterfactuals. These hold that "a counterfactual is true, or assertable, if and only if its antecedent, together with suitable further premises, implies its consequent"; Lewis, *Counterfactuals* (Cambridge: Cambridge University Press, 1973), 65. The "further premises" may include both facts and causal laws, or "lawlike generalizations." For example, the counterfactual "If that match had been struck, it would have lit" is true given the existence of certain laws concerning sulfur, oxygen, friction, and heat, plus certain factual conditions, including a dry match, presence of oxygen, etc. A counterfactual is thus a "condensed or incomplete argument" (J. L. Mackie, "Counterfactuals and Causal Laws," in R. J. Butler, ed., *Analytical Philosophy* [Blackwell: Oxford, 1962], 68). There are other accounts of what makes a counterfactual true (or assertable), based on notions of distance between "possible worlds"; see Lewis.

[15] Van Evera (fn. 7); Jack Snyder, "Civil-Military Relations and the Cult of the Offensive, 1914 and 1984," *International Security* 9 (Summer 1984), 108–46.

[16] Of course, each step of this process—from identifying a sample to interpreting relative importance—is fraught with methodological peril. *Both* strategies, it should be emphasized, are risky.

which often goes under the name "case study."[17] In this instance, careful researchers would make an explicit effort to imagine the prewar world without a cult of the offensive but otherwise similar. They would then construct an argument showing that the outbreak of a general war would have been much less likely in the counterfactual case. Such an argument would depend for its credibility on the principles and historical knowledge used to draw the picture of what would have happened. Stephen Van Evera adopts precisely this strategy to support his cult of the offensive hypothesis, relying at bottom on general principles of rationality.[18] He asks, in essence: How would statesmen have behaved if they had believed that defense rather than offense had the advantage? A reconstruction of what rational actions would have followed from these beliefs yields the conclusion that escalation would have been much less likely in a crisis like that of July 1914 (that is, in the counterfactual case).

The difference between the two means of hypothesis testing would thus appear to be quite stark, and on one level it is. In the counterfactual strategy the analyst supports one causal hypothesis by *invoking others*—laws, regularities, or principles that are taken as having some independent credibility. In the actual case strategy no other principles need to be invoked directly to support the causal hypothesis: only a strength of association across actual cases matters. Indeed, from this vantage point the counterfactual strategy for "empirically" checking a causal hypothesis seems only indirectly empirical, since the confirmation it provides depends principally on other *theories*, which are presumably themselves supported by empirical evidence from actual case comparisons.

On what may be a deeper level, this apparently central difference between the two strategies seems less sharply drawn. As noted, when the actual case strategy is employed in a nonexperimental setting, the validity of a causal interpretation of the results in contingent on the truth of a counterfactual assumption about the other unspecified, unmeasured causes. We must be ready to accept the proposition that had variable X taken values different from those in the sample, no such other causes of the dependent variable would have been systematically different as well. Our confidence that the other causes would not vary with the independent variables depends on our confidence in our theory about what the

[17] I want to suggest that counterfactual reasoning must underlie efforts to infer or assess the relative weights of causes in case studies where the analyst's degrees of freedom in the actual world are negative. In practice, those who use case studies often resort as well to casual comparisons with other actual cases (e.g., "Whereas in many other African countries . . . , in Kenya . . .") and testing multiple implications of a theory; see Donald Campbell, " 'Degrees of Freedom' and the Case Study," *Comparative Political Studies* 8 (July 1975), 178–93.

[18] Van Evera (fn. 7).

other causes are and about how they might be related to the variables being tested explicitly.

Two other contrasts between the counterfactual and actual case strategies should be noted. The first concerns the appraisal of relative causal weight. In the actual case strategy, such appraisals can be carried out in several ways, essentially by contrasting our estimates of the effects of different independent variables.[19] Ultimately, we can do this because we have a sample from which relevant frequencies and magnitudes can be extracted. In the counterfactual strategy, by contrast, we have no concrete frequencies or magnitudes, and the degrees of freedom problem will bite every time we introduce a new variable that may have influenced the particular event to be explained. Explicit justification of claims about relative effects will require *a proliferation of counterfactual cases*.

Suppose, for example, a historian or political scientist wishes to argue that both A and B were causes of event E, but that A was more important than B. The above analysis would suggest that we now need not one but at least two counterfactual scenarios to support this claim. We would need to contrast a counterfactual case where A is present but B absent with one where B is present but A absent, and then invoke general principles and relevant facts to argue that E would have been more likely to have occurred in the first instance.[20]

One might well object that such arguments about what would have happened in multiple counterfactual scenarios will be very imprecise and uncertain. The second contrast between the two strategies relates to this issue of "precision of estimates." In the actual case strategy when N is large, frequencies and magnitudes allow the researcher to get an idea of how much risk attaches to the belief that the true causal effect of a variable is as distinct from the null hypothesis as the results show. In the counterfactual strategy there is no such formal criterion for gauging the risk of error associated with some independent variable. All depends instead on the plausibility of arguments about what would have happened. As will be seen in the example of the debate on the origins of World War I, *arguments about the relative importance of possible causes become arguments about the relative plausibility of different counterfactual scenarios.*

[19] There is, however, more than one meaningful sense to the idea of causal importance in a regression model. See J. Merrill Shanks, "The Importance of Importance" (Working paper, Survey Research Center, University of California, Berkeley, 1982); Christopher Achen, *Interpreting and Using Regression* (Beverly Hills, Calif.: Sage Publications, 1982).

[20] Some philosophers of history working on the problem of how historians can and should attribute causal weightings have proposed similar criteria. See Raymond Martin, "Causes, Conditions, and Causal Importance," *History and Theory* 21 (1982), 53–74, and citations therein.

The analysis in this section bears on current methodological issues in the field in at least two ways. First, scholars in comparative politics and international relations often argue that because statistical methods are inapplicable when we have few cases and many variables, other methods need to be developed to enable sound explanations; these include, among others, the comparative method, structured, focused comparisons, process tracing, and what are sometimes called "qualitative methods." Following the analysis here, we would emphasize that statistical methods are inapplicable in these circumstances for a good reason, namely, the lack of enough cases to support a causal claim. Further, statistical logic implies that assessing a causal claim would require the addition of counterfactual or actual cases. Statistical principles do not simply cease to operate when the number of actual cases dips below twenty or fifteen or ten, creating room for alternative ways of testing causal hypotheses.

Second, researchers should choose between the two strategies of confirmation on the basis of the types of risks they are willing to run. Sometimes the counterfactual claims needed to support a causal inference seem entirely unproblematic. For example, we do not require a formal survey and regression analysis to support the claim that a gunshot through the heart caused the death. Less trivially, a researcher might be skeptical of regression analysis showing no clear relation between domestic political trouble and the initiation of war if it seemed clear from counterfactual reasoning that in a number of cases, domestic problems were a factor impelling the leadership to start a war.[21] Where there are serious problems in identifying a sample, operationalizing and measuring variables, and conceiving of relevant controls, counterfactual argument about one or several cases may be more compelling than a statistical effort.

Indeed, understanding that one can try to explain counterfactual variation in single cases, as well as actual variation across actual cases, may help resolve some of the puzzle over how case studies function methodologically to assess theories and hypotheses.[22] There is a substantial amount of work in political science where the analyst declares an interest in explaining phenomenon X (for example, war, revolution, democracy), chooses a set of cases where X actually occurred, and ends up drawing conclusions about the causes of phenomenon X. Those who engage in large-N analysis tend to view this procedure as totally invalid. Such analysts "sample on their dependent variables"; if they fail to include cases

[21] Cf. Jack Levy, "Domestic Politics and War," *Journal of Interdisciplinary History* 18 (Spring 1988), 653–73.

[22] Another tack on this puzzle is taken by Campbell (fn. 17).

where X does not occur, how can they find causes that differentiate between outcomes? But if we see that each case study proposes causes that selected the actual outcome from a range of possible *counterfactual* outcomes in that case, we see the source of the "not Xs" and the variance such analysts account for. This does not justify the approach—particularly since it is usually performed unconsciously—but it does make some methodological sense of it.

COUNTERFACTUAL ARGUMENT IN PRACTICE

The most controversial point made above is probably that concerning the *necessity* of counterfactual argument for justifying causal claims in small-N settings. The approach is not put forward as simply another option on the menu for small-N practitioners. Rather, I have argued, the point is that when degrees of freedom in the actual world are negative, a causal claim *requires* argument about counterfactual cases for its justification (or addition of other actual cases). This section will consider some examples of how this logical constraint makes itself felt in practice.

One does not find counterfactuals playing central roles in *all* small-N political science research. My impression, after reviewing literature for examples and evidence, is that counterfactuals are most likely to be found performing confirmatory work in case studies where the analyst is explicitly concerned with giving a causal explanation for some event or phenomenon.[23] Of course, case studies may be used for other purposes, such as evaluating the performance of rival theories or simply giving information relevant to various theoretical concerns.[24] In addition, even in what is nominally a case study analysts often employ both strategies of confirmation.

I will first discuss the use of counterfactuals in three examples of N = 1 case studies.[25] Here counterfactual scenarios *must* be developed to support explicit causal claims and to support assertions about relative causal weight. Second, I will consider the role of counterfactuals in sev-

[23] E.g., Youssef Cohen, "Democracy from Above: The Political Origins of Military Dictatorship in Brazil," *World Politics* 40 (October 1987), 30–54; Hyug Baeg Im, "The Rise of Bureaucratic Authoritarianism in South Korea," *World Politics* 39 (January 1987), 231–57; Joanne Gowa, "Hegemons, IOs, and Markets: The Case of the Substitution Account," *International Organization* 38 (Autumn 1984), 661–83.

[24] For examples, see articles in Frederic Deyo, ed., *The Political Economy of the New Asian Industrialism* (Ithaca, N.Y.: Cornell University Press, 1987).

[25] As the preceding discussion should suggest, an N = 1 case study in which causal inferences are drawn is, strictly speaking, impossible, since other counterfactual cases must be invoked to support causal claims. I use N here to refer to the number of cases in the actual world. On the idea of actual versus possible worlds, see Michael Loux, ed., *The Possible and Actual: Readings in the Metaphysics of Modality* (Ithaca, N.Y.: Cornell University Press, 1985).

eral N > 1 examples from comparative politics. Here analysts often make primary use of the actual case strategy when grounding causal assertions, but even then counterfactuals may be needed to justify inferences fully.

N = 1 EXAMPLES

Some of the clearest examples of the importance of counterfactual argument come from research on the causes of World War I. Over the years political scientists and historians have identified an enormous collection of possible factors,[26] which are typically argued to be causes on the following grounds: If cause X had not been present, the war either would not have occurred or would have been much less likely to have occurred. Thus, in arguing the causal importance of misperceptions in 1914, Robert Jervis writes, "Had the participants realized not only that the first offensive would not end the war, but also that the fighting would last for four punishing years, they might well have held back."[27] Note that Jervis relies on a rationality principle (sensitivity to war costs) to make credible the causal inference drawn from the counterfactual proposition.

On similar grounds, Van Evera has developed the thesis that a cult of the offensive was a major cause of World War I.[28] Further, he argues that military and civilian tendencies to glorify the offensive had the effect of "feeding or magnifying a wide range of secondary dangers" that other analysts thought were independent or unrelated causes. To establish this, Van Evera discusses the secondary dangers one by one, arguing in each case that had the cult of the offensive not been present, the secondary cause *would not have* operated with as much (or any) force. His conclusion nicely summarizes these counterfactual arguments. Throughout, Van Evera relies primarily on implicit rationality principles: he supposes leaders had different beliefs and then draws conclusions about appropriate or rational behavior given such beliefs.[29]

[26] These include, but are not limited to, nationalism, imperialism, capitalism, social Darwinism, a fatalistic intellectual mood, the balance of power system, population growth, differential industrialization, a power transition, long cycles, tight alliances, multipolarity, misperceptions, psychological pathologies, leader personalities, essentially aggressive German intent, military doctrine (i.e., the cult of the offensive), military organization, diplomatic errors, the Russian mobilization, the archduke's assassination, and the outcomes of recent crises.

[27] Jervis, "War and Misperception," *Journal of Interdisciplinary History* 18 (Spring 1988), 684.

[28] Van Evera (fn. 7).

[29] I should note that rationality principles are not the only ones that might be used to limn counterfactual scenarios. One might argue, for example, that had some independent variable been different, a key actor would have ignored it due to cognitive dissonance or wishful thinking.

Even so, the frequent use of rationality principles to sketch counterfactual scenarios should not be surprising. The counterfactual strategy is often used by analysts explaining an outcome

> The consequences of the cult of the offensive are illuminated by imagining the politics of 1914 *had* European leaders recognized the actual power of the defense.... All European states *would have been* less tempted to mobilize first, and each *could have* tolerated more preparations by adversaries before mobilizing themselves, so the spiral of mobilization and counter-mobilization *would have* operated more slowly, if at all. If armies mobilized, they *might have* rushed to defend their own trenches and fortifications, instead of crossing frontiers, divorcing mobilization from war. Mobilizations *could more easily have been* confined to single frontiers, localizing the crisis. Britain *could more easily have* warned the Germans and restrained the Russians, and all statesmen *could more easily have* recovered and reversed mistakes made in haste or on false information. Thus the logic that led Germany to provoke the 1914 crisis *would have been* undermined, and the chain reaction by which the war spread outward from the Balkans *would have been* very improbable. In all likelihood, the Austro-Serbian conflict *would have been* a minor and soon-forgotten disturbance on the periphery of European politics.[30]

The use of counterfactuals is explicit and clear in Van Evera's analysis because he is methodologically self-conscious about providing a causal explanation. This is less true of much historical scholarship on the causes of World War I, where the key counterfactual propositions are often left implicit or underdeveloped.

Explicit treatment of counterfactual cases may in turn have the advantage of sharpening substantive debates. In the example at hand Scott Sagan has offered some important qualifications to the arguments advanced by Van Evera and Jack Snyder. He argues, among other things, that Van Evera and Snyder "have overlooked the negative consequences that *would have resulted* if the great powers had adopted purely defensive military doctrines."[31] He takes issue, in other words, with Van Evera's counterfactual scenario. Sagan holds that the offensive doctrines of the major European powers were rationally chosen to provide extended deterrence to key strategic allies and were not simply or solely the result of the biases of military organizations. He suggests that defensive doctrines might have left states unable to offer credible threats on behalf of their allies, possibly increasing the chances for smaller wars that would have

---

as the result of human choices. This entails saying why other possible choices were not seen as desirable by the actors. In game-theoretic terms, analysts using the counterfactual strategy are often describing why some particular set of choices was an equilibrium (or, at least, rationalizable) strategy in the "game" faced by the actors. On Nash equilibrium versus rationalizability as game-theoretic solution concepts, see B. Douglas Bernheim, "Rationalizable Strategic Behavior," *Econometrica* 52 (1984), 1007–28.

[30] Van Evera (fn. 7), 105 (emphasis added).

[31] Scott Sagan, "1914 Revisited," *International Security* 2 (Fall 1986), 151–75, at 159 (emphasis added).

altered the balance of power against them (for example, Germany loses Austria to Russia, or Russia loses France to Germany). Sagan's argument contains excellent examples of the use of general principles and specific historical knowledge to support counterfactual scenarios, as well as some clever uses of the actual case strategy to refute counterfactual claims with which he disagrees.[32]

Another good example of counterfactual analysis in international relations research focuses on the *non*occurrence of an important phenomenon, namely, that there has not been a war among major powers since 1945. This outcome might be explained by any of the following causes, according to different theories: bipolarity, the presence of nuclear weapons, successful balance of power politics, or the obsolescence of major war due to the "Hollandization" of the great powers.[33] If either nuclear weapons or Hollandization were in fact the true or major cause of postwar military stability, then we cannot hope to employ the actual case strategy to check this, since neither variable varied much before 1945.[34] With John Mueller, who has recently argued the case for Hollandization against the more widely accepted nuclear weapons thesis, we would be compelled to argue about what would have happened had nuclear weapons not been invented and amassed in this period.[35] As Mueller puts it:

> The postwar world *might well have* turned out much the same even in the absence of nuclear weapons. Without them, world war *would have been* discouraged by the memory of World War II, by superpower contentment with the postwar status quo, by the nature of Soviet ideology, and by the fear of escalation [to conventional war].[36]

Mueller proceeds to argue the counterfactual case for each of these "independent variables" favoring postwar stability. Though he does not deny that nuclear weapons may have had some damping effect on poten-

[32] See also Snyder's response to Sagan's critique and Sagan's reply, *International Security* 9 (Winter 1986–87), 187–98. Their discussion is carried out largely in the realm of the counterfactual (e.g., what was the *probability* that the Schlieffen plan would work).

[33] The Hollandization thesis is developed by John Mueller in *Retreat from Doomsday* (New York: Basic Books, 1989), where he argues that gradual changes in the government and societies of advanced industrial states have made them more peaceable in their external affairs. For a review of arguments on the causes of the long peace, see John Lewis Gaddis, *The Long Peace* (Oxford: Oxford University Press, 1987), chap. 8.

[34] Depending on how one counts the "poles," neither does bipolarity; see Waltz (fn. 1).

[35] To assess the question of *relative* importance, we would also need to ask about what would have happened if nuclear weapons existed but Hollandization did not. Mueller does not explore this second counterfactual scenario explicitly. To hold that Hollandization has been the more important cause, he would need to argue that postwar states lacking the key Hollandization attributes might not have been deterred from fighting a major war, despite nuclear weapons.

[36] Mueller, "The Essential Irrelevance of Nuclear Weapons: Stability in the Postwar World," *International Security* 13 (Fall 1988), 55–79, at 56 (emphasis added).

tial escalation, he holds that their causal effect has been redundant, du to the combined impact of the other variables. The claim about the coun terfactual case—the postwar world with no nuclear weapons and no ma jor war—is supported by some specific historical detail (for example characteristics of Soviet ideology) and by at least one general principle "Wars are not begun out of casual caprice or idle fancy, but because on country or another decides that it can profit from (not simply win) th war—the combination of risk, gain, and cost appears preferable to peace."[37] Taking this as either a theoretically plausible or an empiricall confirmed regularity, Mueller suggests that even disregarding the adde costs posed by nuclear weapons, the costs of conventional war in thes years would have been enough to deter the U.S. and Soviets from a ho war.[38]

A final example of the counterfactual strategy as used in an N = 1 case study comes from work on the breakdown of democratic regimes.[39] Alfred Stepan's explanation for the 1964 military takeover in Brazil illustrates a fairly common way that counterfactuals are employed in comparative politics and international relations case studies.[40] Stepan proposes that the actual outcome—the military coup—was made possible by the operation of certain social, economic, and ideological "macropolitical" factors but that these did not make the coup "inevitable." "There remained a small margin of maneuverability within which the process of increasing democratization and participation could have been expanded."[41] Stepan is here defining the *range of counterfactual variation* that he wishes to explain. Brazil in 1964 could have seen a democratic outcome but did not. The micropolitical factors that reduced the "margin of maneuverability" and selected the authoritarian outcome from the range of possibilities are then assigned causal status above that of the

[37] Ibid., 68–69.

[38] The fortunate absence of actual cases of nuclear conflict has led a number of historians and political scientists to reflect on the role of counterfactuals in nuclear history. See John Lewis Gaddis, "Nuclear Weapons and International Systemic Stability," American Academy of Arts and Sciences Occasional Paper No. 2 (Cambridge: AAAS, 1990). This paper was prepared for an AAAS workshop entitled "Nuclear History and the Use of Counterfactuals." In a different vein, Richard Ned Lebow and Janice Gross Stein ("Beyond Deterrence," *Journal of Social Issues* 43 [Winter 1987], 3–71) have briefly discussed the role of counterfactuals in defining a sample of cases of successful deterrence.

[39] Juan Linz and Alfred Stepan, eds., *The Breakdown of Democratic Regimes* (Baltimore, Md.: Johns Hopkins University Press, 1978).

[40] Stepan, "Political Leadership and Regime Breakdown: Brazil," ibid. For other examples, see Richard Smoke, *War: Controlling Escalation* (Cambridge: Harvard University Press, 1977), and citations in fn. 23.

[41] Stepan (fn. 40), 134, and see also 120.

macropolitical factors, which, in Stepan's view, were not "sufficient" themselves to determine the result.[42]

In his historical treatment of the events leading up to the coup, Stepan identifies political strategy choices by the incumbent president, João Goulart, as the key micropolitical causes of the breakdown of the democratic regime. In an atmosphere of political stalemate, Goulart lost important military and middle-class allies by proposing major economic and constitutional reforms and bidding for the support of the left to back them. But still, "as late as twelve days after [the declaration of these reforms] no 'winning coalition' existed to overthrow Goulart." A naval mutiny by lower-level officers and sailors then occurred, forcing Goulart to choose between alienating either the mutineers or the higher-level officers, who saw the mutiny as a major "threat to the principle of military discipline."[43] His decision to be lenient with the mutineers had the unforeseen effect of galvanizing high-level military support for a coup.

These two key political choices are posed as causes of the breakdown of the regime on counterfactual grounds: had Goulart chosen different strategies, the analysis suggests, a coup might not have occurred. The counterfactual contrasting case is justified by reference to specific historical detail (evidence that the military was divided and generally not supportive of direct military rule before the choices were made) and to general principles (for example, the proposition that plotters do not act unless they expect sufficient support, or lack of resistance, from other key actors).

Though Stepan is much more careful and explicit about his counterfactual comparison than is often the case, I would argue that his analysis only goes part of the way to justify his causal claims, essentially because he does not spell out the counterfactual scenario in quite enough detail. Goulart's *reasons* for choosing the left-oriented, constitutional reform strategy, and thus in a sense the deeper causes of the takeover, are left unclear. Stepan seems to suggest that Goulart's destabilizing move leftward was more a function of his personality and aspirations than it was of the untenability of other alternatives. But we need more careful speculation about what would have happened had he instead stayed with his divided and indecisive coalition. If in the longer run his position was simply impossible—that is, if no civilian leader could govern given the political stalemate under existing institutional arrangements—then the

[42] The distinction is similar to that between underlying causes and specific or proximate causes—a framework often used by historians.

[43] Stepan (fn. 40), 129 and 130.

macropolitical factors would seem to gain in importance as causal factors.[44]

N > 1 but Still Small-N Examples

Researchers with more than one actual case are not logically compelled to use the counterfactual strategy to justify a causal claim, as long as they do not have more independent variables than cases (less one), or two or more independent variables that vary together (perfect multicollinearity). Roughly speaking, these conditions ensure that regression estimates can be derived, and they are usually met with ease in large-N research projects. Quite often, however, researchers in comparative politics and international relations work with few cases and many variables.[45] There are sometimes opportunities in this intermediate range to employ the actual case strategy, but the application of statistical methods either would fail to yield estimates of causal effects or would yield wildly imprecise estimates. In these circumstances, I would argue, one typically finds a mixing of the actual and counterfactual case strategies, with each used to make the other more credible. In good large-N research, the credibility of causal effect estimates derives in the first instance from ample degrees of freedom. Given the theoretical assumptions supporting a causal interpretation, causal claims are empirically supported by regularities of association, in Humean fashion. With an N between 2 and (say) 15, however, the regularity justification is weaker and may need support from more detailed treatments of individual cases. Readers may want to know not only that the proposed causes correctly partition outcomes across the few actual cases, but that in each case the proposed causes indeed *produced* the effects attributed to them. In such efforts one can find examples where researchers resort implicitly or explicitly to the counterfactual strategy and examples where they would have to use it to defend their causal claims.

One common methodological practice in comparative politics and international relations work could be called the "loading up of explanatory factors." The researcher lists several causes for the phenomenon being explained, all of which were present in the cases where the phenomenon occurred. In formal terms, the researcher has a multicollinearity problem. In such instances, counterfactual arguments would be necessary to

[44] In a current project, Stepan uses explicit counterfactual analysis to assess the impact of presidential as opposed to parliamentary systems on democratic regime breakdown in South America and Southern Europe.

[45] Arend Lijphart, "Comparative Politics and the Comparative Method," *American Political Science Review* 65 (September 1971), 682–93.

support the claim that *any one* of the proposed conditions has a causal effect.

Consider, for example, one of Stepan's arguments in his work on Peru.[46] His goal is to explain the success or failure of attempts to install corporatist political arrangements in Latin American states. He identifies five independent variables and gives general hypotheses linking each to the likelihood of success or failure. These five variables are then shown to discriminate between actual cases of success and failure in the following sense: where they were all basically favorable to corporatist installation, installation succeeded; where they were basically unfavorable, it did not. The difficulty here is that without counterfactual argument, we cannot decide which of these variables mattered—whether at all or how much. It could be, for example, that only one or two of these variables is really critical and that the rest are totally irrelevant. There are only two ways to decide: (1) find new actual cases where one explanatory factor is present but others are not, or (2) argue counterfactually that the removal of any one of the variables would have damaged chances for corporatist success in the actual cases we have. The tendency to "load up" explanatory factors is quite common. For instance, Barrington Moore's landmark work contains many examples of this practice, such as the list of five "main conditions that have apparently been most important for the development of democracy.[47]

Counterfactuals may also come into play in what is nominally actual case work when analysts use historical treatments of particular cases to make credible claims based on actual case associations. For example, Theda Skocpol identifies three key variables that differentiate her "positive cases" of social revolution (1789 France, 1917 Russia, and 1911–49 China) from actual cases in which social revolutions did not occur (for example, Meiji Japan, seventeenth-century England, 1807 and 1848 Prussia, Russia after the Crimean War and in 1905, early-eighteenth-century France).[48] Rather than simply stating the values of the independent variables for the different cases and showing that they differentiate between outcomes, Skocpol undertakes moderately extensive historical treatments of each positive case, detailing how the independent variables she identifies produced social revolution in each one. Though Skocpol makes

[46] Stepan, *The State and Society: Peru in Comparative Perspective* (Princeton: Princeton University Press, 1978).

[47] Moore (fn. 1), 430. See also Alexander L. George et al., *The Limits of Coercive Diplomacy* (Boston: Little Brown, 1971), 227.

[48] Skocpol (fn. 4). Only four of these "negative cases" are treated explicitly and at length, though Skocpol is well aware that others mentioned are used in the same fashion.

frequent use of the actual case strategy within historical treatments,[49] her approach is broadly similar to that of Stepan in the Brazil example, in that the independent variables are shown to select out certain historical actualities from a range of possibilities.

Gregory Luebbert's exemplary use of the actual case approach in a small-N setting and provides some final examples of how the counterfactual strategy may be employed in such analyses. Luebbert first identifies two independent variables that perfectly partition his fourteen actual cases of European interwar regime types. "Pluralist democracies" occurred only in countries where liberal parties gained dominance before World War I. Regime type elsewhere was determined by which party successfully formed a coalition with the rural "middle peasants": if it was socialists, then "corporatist democracy" resulted; if liberals, then "traditional dictatorship"; if neither, then fascism. Luebbert does not dwell on the causal links between coalition membership and regime type. In justifying the causal argument that the effects of World War I made pluralist democracy along the lines of Britain and France improbable elsewhere, he *does* explore the following counterfactual: "In . . . Italy, Norway, and Sweden, another generation of peace *might have resulted* in pluralist democratic regimes."[50] But on the whole he is content to let the perfect association and the intuitively acceptable idea that coalition members determine the policy regime support the causal claim.

Instead, Luebbert turns his analytic attention to identifying "the conditions that produced each of these coalitions."[51] By implication, these will be the final or deeper causes of regime type. This effort is marked by both actual case comparisons and implicit counterfactual arguments. For examples of the latter, consider Luebbert's explanation of why socialist parties allied with middle peasants rather than with the agricultural proletariat in Norway and Denmark (thus yielding corporatist democracy). In Norway, he notes, there were few landless laborers, so they were not a tempting group for socialists to mobilize. Rather than correlating size of rural worker populations with socialist mobilization efforts across several actual cases, Luebbert simply appeals to a rationality principle that would support the appropriate counterfactual argument.[52] In Denmark the socialists could not mobilize what was a much larger ag-

[49] Ibid., e.g., 63.

[50] Luebbert (fn. 8), 457–58 (emphasis added).

[51] Ibid., 452.

[52] The rationality principle is: Parties desirous of electoral success will seek partners that can carry many votes with them. The implicit counterfactual argument is: If there had been many landless laborers in Norway, the socialists might have sought to form a coalition with them, and fascism might have resulted.

icultural labor force because "this population had already been heavily nobilized by another party."[53] Instead, they mobilized the middle peasınts, leading to the corporatist coalition. The implicit counterfactual is: f the agricultural labor force had not already been mobilized, then it night have been mobilized by the socialists, and fascism rather than cor›oratism would have resulted. Thus, a particular fact about Danish prevar politics becomes an ultimate cause of corporatism rather than dicatorship or fascism in that country.

Counterfactuals and Causation: Two Theoretical Issues with Practical Implications

The proposition that a cause of a particular historical event may be esablished by imagining the effect of its (counterfactual) absence has been nade before. In what remains one of the best essays on the topic, Weber rgued vigorously for recognition of the link between causal explanation nd counterfactuals in historical research.

> [The question of] what might have happened if, for example, Bismarck had not decided to make war [in 1866] is by no means an "idle" one [contrary to the view of historian Eduard Meyer]. It does indeed bear on something decisive for the historical moulding of reality, namely, on what causal *significance* is properly attributed to this individual decision in the context of the totality of infinitely numerous "factors" . . .[54]

Since Weber has been a methodological guru for generations of socilogists and political scientists, it is somewhat surprising that this particılar essay has been so little discussed and explicitly applied. In recent ears the only serious and sustained debate on the role of counterfactuals utside of philosophy took place among historians (and without refernce to Weber), in their discussions of the use of counterfactuals by some ractitioners of the "new economic history."[55] The only political scientist know of who has examined the topic at length is Elster, particularly in is *Logic and Society*.[56] He presents a novel "branching worlds" theory

[53] Luebbert (fn. 8), 466.

[54] Weber (fn. 10), 164 (emphasis in original).

[55] Robert Fogel, *Railroads and American Economic Growth* (Baltimore, Md.: Johns Hopkins Iniversity Press, 1964); McClelland (fn. 9); Gould (fn. 9); Fritz Redlich, " 'New' and Traitional Approaches to Economic History and Their Interdependence," *Journal of Economic Iistory* 25 (1965), 480–95; and T. A. Climo and P. G. A Howells, "Possible Worlds in Hisrical Explanation," *History and Theory* 15 (1976), 1–20. Fischer (fn. 9) lists further refer›ces.

[56] Elster (fn. 10). See also Elster, *Explaining Technical Change* (Cambridge: Cambridge niversity Press, 1983), chap. 1; idem, "Reply to Comments," *Inquiry* 23 (June 1980), 213–2; Steven Lukes, "Elster on Counterfactuals," *Inquiry* 23 (June 1980), 145–55; Brian Barry,

for assessing the validity of counterfactual propositions and uses it to analyze some examples in economic history. Outside the social sciences, of course, analytic philosophers have been writing about counterfactuals and causation for years.[57] While much of this literature (for example, that concerned with the metaphysics of modality)[58] would seem largely irrelevant to social scientists, some recent work on counterfactuals and explanation by philosophers of history has practical value and probably deserves greater attention.[59]

Scholars who have dealt with counterfactuals have often expressed dismay, doubt, and bewilderment at the sorts of logical and philosophical problems such propositions seem to entail. This section briefly introduces two problems that seem particularly bothersome to social scientists and historians.

The first is sometimes referred to as the "Cleopatra's Nose Problem."[60] According to Pascal, if Cleopatra's nose had been shorter, Antony might not have been so infatuated, and the course of Western history might have been different. Does this imply that the gene controlling the length of Cleopatra's nose was a cause of World War I? More generally, if we believe that an event A satisfies

(P2) If A had not occurred, B would not have occurred,

then are we committed to saying that A was a cause of B?

This is not just an idle question. As we have seen, social scientists often argue that A was a cause of B on precisely these grounds—that had A not occurred, B might not have occurred. How do we distinguish be-

---

"Superfox," *Political Studies* 28 (1980), 139–43. Political scientists have broached issues raised by counterfactuals in a variety of places. See, for example, Alexander George and Timothy McKeown, "Case Studies and Theories of Organizational Decision Making," in Robert Coulam and Richard Smith, eds.: *Advances in Information Processing in Organizations* (Greenwich, Conn.: JAI Press, 1985), 2:33–34; Charles Ragin, *The Comparative Method* (Berkeley: University of California Press, 1987), 39; Donald Moon, "The Logic of Political Inquiry: A Synthesis of Opposed Perspectives," in Nelson Polsby and Fred Greenstein, eds., *Handbook of Political Science* (Reading, Mass.: Addison-Wesley, 1975), 1:131–228.

[57] Nelson Goodman, "The Problem of Counterfactual Conditionals," *Journal of Philosophy* 44 (1947), 113–38, reprinted in his *Fact Fiction and Forecast* (Cambridge: Harvard University Press, 1983); Lewis (fn. 14); Ernest Sosa, ed., *Causation and Conditionals* (Oxford: Oxford University Press, 1975). Part of the philosophical interest in counterfactuals arises from their bearing on key issues in the philosophy of science. See Frederick Suppes, "The Search for Philosophic Understanding of Scientific Theories," in Suppes, ed., *The Structure of Scientific Theories* (Urbana: University of Illinois Press, 1977), 3–232, at 36–45, and references cited there; Ernest Nagel, *The Structure of Science* (New York: Harcourt, Brace, and World, 1961). Nagel (chap. 15) also saw that counterfactuals play a key role in historical explanation.

[58] Loux (fn. 25).

[59] Martin (fn. 20). See also Martin, "Beyond Positivism: A Research Program for Philosophy of History," *Philosophy of Science* 48 (1981), 112–21; and idem, "Singular Causal Explanation," *Theory and Decision* 2 (1972), 221–37.

[60] Edward Hallet Carr, *What Is History?* (New York: Knopf, 1962); Gaddis (fn. 38).

tween the infinity of particular factors that would have precluded or reduced the likelihood of some interesting event had they *not* taken place?

One approach would be to hold that causality should not be defined in terms of counterfactuals like P2, that "A satisfies P2" does not imply that A is a cause of B. Consider the view that a cause is something that *produces* its effect whenever (or usually when) it occurs. The cult of the offensive can be understood to have produced World War I in this sense, but Cleopatra's nose really cannot. This strategy amounts to accepting a regularity theory of causation.[61] Accidental happenings that help lead to specific events are not "causes" but only "conditions"; conditions of particular events that generalize or could regularly produce the effect are labeled causes. The distinction between causes and conditions could conceivably be a useful one for political scientists engaged in small-N work, and particularly for case studies.[62] The distinction can do violence to common sense and ordinary usage, however, as in the following: the unlucky person's death was not "caused" by the falling shingle, it was caused by skull fracture; the shingle was only a "condition."

Another approach would be more lenient with certain accidental happenings. We could argue that the length of Cleopatra's nose did not make World War I any more likely than myriad other possible worlds that could have followed, whereas the presence of the cult of the offensive did significantly "select out" the particular outcome that was World War I. That is, the same theoretical argument holding that the probability of World War I conditional on Cleopatra's nose being shorter was zero implies as well that the probability of World War I conditional on her nose being as it was must have been *almost* zero. On this account, an accidental (or "random") happening—say, a monkey bite leads to the death of a king, whose replacement begins a war[63]—could qualify as a "cause" of a particular event. The important point is that in both accounts events that satisfy P2 are not necessarily causes of the phenomenon being explained. Though counterfactuals like P2 might be explored to lend credence to a causal claim, a cause does more than just satisfy P2.[64] Both suggestions could have practical value for political scientists arguing causality on counterfactual grounds.

[61] On these, see Tom Beauchamp and Alexander Rosenberg, *Hume and the Problem of Causation* (Oxford: Oxford University Press, 1981).

[62] For example, Luebbert (fn. 8) might have distinguished more carefully between the *conditions* prevailing in particular countries that allowed the *causes* of regime type—coalition membership—to operate as they did. On related philosophical distinctions between causes and conditions, see J. L. Mackie, "Causes and Conditions," in Sosa (fn. 57), 15–38; and Martin (fn. 59, 1981, 1972).

[63] Carr (fn. 60), citing Churchill.

[64] A third suggestion for resolving this problem would be to add a condition of temporal

The second vexing problem for scholars concerns what Elster calls "legitimacy" of a counterfactual proposition.[65] Elster argues that a counterfactual thought experiment undertaken to assess a causal hypothesis is not legitimate if we have a theory saying that the counterfactual antecedent could not have happened. Suppose we wish to learn the effect of the railroad on the growth of GNP in nineteenth-century America and that we attempt to do so by imagining the last century without the railroad.[66] Elster thinks it nonsensical to speculate whether the internal combustion engine would have been invented earlier than it was (in the counterfactual nineteenth-century America without railroads), since an answer would require a theory of technical change strong enough to make the original counterfactual proposition implausible. If we could predict whether the gas engine would have been invented earlier, surely we would also have a theory showing that the railroads "had to be" invented when they were. Elster calls this "the scissors problem,"[67] or "the unimportance of inevitable":[68] the better our theories, the more things we know "had to occur" as they did, and thus the fewer counterfactuals we can legitimately assert.

In social science practice this problem often appears in the following guise. On the basis of actual case comparisons, a comparativist claims that C caused E in country X, suggesting that if C had been different, the outcome in country X might have been more like the outcome in country Y. A specialist on country X criticizes this as absurd, arguing that due to a complex of historical and cultural factors particular to country X, C could not have been different.

The insistence that counterfactual propositions be "legitimate" may confound, or at least obscure, two distinct problems.[69] The real issue is not legitimacy, if this means that in our counterfactuals we cannot legitimately vary causes that had to occur as they did. Whether event C had to occur has no direct bearing on its causal status with respect to E. A variable may help explain one outcome and still itself be explained by the action of other variables. In large-N work this pattern is commonly found in structural equations models, in which a dependent variable in one equation may be an independent variable in another equation.

---

or causal proximity to P2; that is, A is a cause of B if P2 is true and A precedes B by a relatively short time period, or if the causal chain is not too long. But this raises the problem of how long?

[65] Elster (fn. 10).

[66] Fogel (fn. 55).

[67] Elster (fn. 56, 1983), 38.

[68] Elster (fn. 10), 185.

[69] For related criticisms of Elster's notion of counterfactual legitimacy, see Barry (fn. 56); and Lukes (fn. 56).

The key issue, it seems to me, is what philosopher Nelson Goodman called "cotenability." In Goodman's account, a counterfactual assertion is judged true if (1) the counterfactual antecedent, when joined with appropriate theories and facts, implies the consequent; and (2) the counterfactual antecedent is "cotenable" with the facts or "initial conditions" used to draw the inference, meaning that *if the antecedent had actually occurred, the initial conditions could also have occurred.*[70] Thus, in his critique of the cult-of-the-offensive hypothesis, Sagan in effect argues hat supposing the absence of the cult is not cotenable with supposing a 914 otherwise identical to the actual 1914. As noted, he suggests that eliefs in defensive superiority would have created a different strategic roblem for state leaders, one that could also have produced war.

The question, then, is not whether a factor had to occur but whether arying the factor implies changing other factors that *also* would have iaterially affected the outcome.[71] It is not appropriate to criticize a coun-rfactual argument by saying that the antecedent could not have oc-irred. Rather, we need an explicit argument saying that if the antece-:nt had been the case, other changes would be required in the ounterfactual scenario that would have affected the outcome in a differ-it way.

An obvious methodological prescription follows: analysts using the rategy of counterfactual argument should pay close attention to hether their counterfactual suppositions are cotenable with the facts id theories used to draw the causal inferences they make. This is per-ips a more precise statement of what it means to make a counterfactual gument plausible. I expect that in practice, the cotenability require-ent will be more plausibly satisfied for small causes, such as specific licy decisions, than for big causes, such as nationalism, imperialism, or cult of the offensive. History often provides evidence that leaders con-lered several possible choices at certain junctures, and in some in-inces it may be feasible to imagine a different choice without changing her major influences on the outcome in question. The fewer the

[70] Goodman (fn. 57, 1983), 15–17. See also references in fn. 14. Goodman points out that it quite problematic to use a counterfactual to define general truth conditions for counterfac-ils. See Mackie (fn. 14) for a possible way around this problem (which at any rate may be greater interest to philosophers than to political scientists).
I should note that Elster (fn. 10) is well aware of the issue of cotenability, which he refers as "compossibility" (p. 177) and also "compatibility" (p. 183). Indeed, his "branching rlds" theory for assessing the truth of counterfactuals can be seen as a suggestion for essing cotenability.

[71] Note the similarity of the cotenability condition to PI, the key assumption justifying a isal interpretation of regression coefficients derived from quasi-experimental data. The eness underscores the point that quasi experiments and the counterfactual strategy share iance on counterfactual suppositions.

changes from the actual world required by a counterfactual suppositi the easier it will be to draw and support causal inferences, and the mc defensible they will be.[72]

Conclusion

Counterfactuals and the counterfactual strategy of hypothesis testin play an important but often unacknowledged and underdeveloped rol in the efforts of political scientists to assess causal hypotheses. I have trie to show that any *non*experimental research that makes causal claims, b it of the large-N or small-N variety, must confront counterfactuals in th form of key assumptions or in the use of hypothetical comparison cases Particularly in small-N research, the common condition of too man variables and too few cases makes counterfactual thought experiments a necessary means for serious justification of causal claims. I close with tw simple suggestions for analysts evaluating causal claims via counterfac tual argument rather than via regularities of association in a sample o actual cases.

First, small-N analysts could strengthen (or simply specify) thei causal arguments by being *explicit* about the counterfactual scenario needed to support their hypotheses. Quite commonly, researchers ir comparative politics and international relations assert that their dependent variable is X, where X is some *particular* event or phenomenon. X might be the failure of the U.S. to play the role of international hegemor between the world wars, a change in the nuclear proliferation regime the dominance of the Liberal Democratic Party in Japan, or the collapse of communism in Eastern Europe. Analysts explaining such events neec to understand that *none* of these are variables. They become *values* ol variables if alternative, counterfactual scenarios are identified or if actual cases, some of which differ in outcome, are added to the analysis.[73] If, for whatever reason, one is reluctant to add actual cases, then it is essential to make explicit what might have happened if a possible cause had varied. Counterfactual comparison cases need not be exhaustively detailed—just specified—so the reader knows what variation the theory or hypothesis proposes to explain.

The second key step is making inferences drawn from a counterfactual comparison *defensible*. Sometimes the argument implied by a coun-

[72] This suggestion is influenced by examples provided by McGeorge Bundy (fn. 7), and by Gaddis's discussion of them (fn. 38).

[73] The point that a variable is distinct from any particular realization of it should be obvious but is sometimes missed. The point that the variance explained might be defined across actual *or* counterfactual cases is rarely seen.

terfactual proposition is quite obvious and requires *little or no unpacking.*[74] *Other times, particularly when the hypothesis is evaluated against* other hypotheses, analysts should make clear what arguments support it nd how they do so. The analyst needs to ask whether the causal inference does indeed follow from the theories and historical facts used to sketch the comparison case and then, whether the counterfactual proposition is cotenable with the counterfactual scenario. Cotenability requires that if the counterfactual assertion had been true (for example, if there had been no cult of the offensive), nothing else would also have been different in a way that would have materially affected the outcome.

Of course, we can never be certain about what else would have been different if C had been different. But perhaps we can venture arguments that can be judged more credible or less credible, depending on our use of historical detail and theories about the way people behave. I should emphasize that I am not suggesting that there can be a special methodology for determining precisely what would have happened. Nor am I advocating a new methodology to rival established approaches to hypothesis testing. Rather, the intent has been, first, to show that counterfactuals cannot be avoided in nonexperimental hypothesis testing, then, to explicate their roles, and finally, to recommend that when political scientists use counterfactuals, they do so explicitly and carefully.

[74] For example, "if that match had been struck, it probably would have lit" will not be controversial in most circumstances. Neither are the counterfactual arguments implied by Luebbert (fn. 8) on why the fascist coalition did not develop in Norway and Denmark.

SOVIET FEDERALISM AND ETHNIC MOBILIZATION

By PHILIP G. ROEDER*

A CENTRAL element of the Soviet developmental strategy was the creation of political institutions that expanded the control of the regime over the processes of social mobilization associated with modernization. This strategy was noteworthy for providing a considerable measure of interethnic peace as the Soviet regime began the process of industrialization. And yet seven decades after the Soviet regime assumed power, with the industrialization of the economy and urbanization of society well under way, this developmental strategy instead fuels a divisive and destructive ethnopolitics.

The Soviet developmental strategy seems to turn around a pattern familiar in the Western developmental experience. As Ernest Gellner notes: "The age of transition to industrialism was bound" also to be "an age of nationalism."[1] But the Soviet strategy delayed the political reckoning with the "age of nationalism" to a much later stage of industrialization. In the short term this was a prudent means to avoid the simultaneous crises that can overtax the capabilities of a new polity: the Soviet regime did not confront a crisis of identity as it sought to build the foundations of Soviet power and initiate the economic transformation of society.[2] This strategy nonetheless contained the roots of its own longer-term dysfunction and in the past three and a half decades has given rise to new ethnic assertiveness and protest.[3]

The Western experience with peripheral nationalism has differed more significantly from the Soviet pattern in a second respect. In the West—regardless of macroeconomic conditions that occasion the rise of peripheral nationalisms—it has been most pronounced among the less

* For their comments on earlier drafts, I thank my colleagues and good friends—Deborah Avant, Anthony Brunello, Ellen Comisso, Patrick Drinan, Arend Lijphart, Richard Little, Debra Rosenthal, Gershon Shafir, Susan Shirk, Tracy Strong, and Michael Tierney.

[1] Gellner, *Nations and Nationalism* (Ithaca, N.Y.: Cornell University Press, 1983), 40.

[2] Sidney Verba, "Sequences and Development," in Leonard Binder et al., *Crises and Sequences in Political Development* (Princeton: Princeton University Press, 1971), 283–316; Dankwart A. Rustow, *A World of Nations* (Washington, D.C.: Brookings, 1967), 120–32.

[3] I make a parallel argument concerning the relationship between the Soviet developmental strategy and political participation; see Philip G. Roeder, "Modernization and Participation in the Leninist Developmental Strategy," *American Political Science Review* 83 (September 1989), 859–84.

World Politics 43 (January 1991), 196–232

dvantaged. In nineteenth-century Western Europe, according to Joseph Rudolph and Robert Thompson, "the most common causal element giving rise to the urge for autonomy" was the aggravation of a peripheral ethnic group's "marginality in, or exploitation by, the state system to which it belongs."[4] During periods of economic prosperity in the more recent rise of peripheral nationalisms, Peter Gourevitch has found that those most likely to support ethnic political movements are disadvantaged ethnic groups drawn by the opportunity or promise of expanded resources.[5]

In the Soviet Union the rise of ethnopolitics has been most significant in the Caucasian and Baltic republics. It is there that local leaders have pressed the most ambitious legislative agendas for change (see Table 1) and that citizens have mounted the largest and most frequent demonstrations (Table 2). Yet, as Figure 1 shows, these nationalities—particularly the Armenians, Georgians, and Estonians—are among the most successful ethnic groups in terms of educational and occupational attainment, in many instances reporting rates even higher than those for the numerically predominant Russian population. Even in the area of Party membership, Georgians report far higher rates among the adult population than do the Russians, and the Armenians report rates above the average for all nationalities. Far less inclined to mount this form of political action in recent years have been the least advantaged nationalities, such as those in Central Asia. Thus, it is the nationalities with the highest levels of educational, occupational, and often political attainment, rather than the disadvantaged or marginal ones, that have advanced the most ambitious agendas for change and engaged in the most extensive protest.

The new Soviet ethnopolitics is structured by the federalism of nominally autonomous ethnic homelands. Appreciating the strategic value of organizational weapons, political entrepreneurship, and mobilizational resources, the architects of the Soviet regime came to understand that federal institutions could expand their control over the politicization of ethnicity. Within each homeland the regime created a cadre of party and state officials drawn from the indigenous ethnic group but dependent

[4] Rudolph and Thompson, "Ethnoterritorial Movements and the Policy Process: Accommodating Nationalist Demands in the Developed World," *Comparative Politics* 17 (April 1985), 292. See also Ernest Gellner, *Thought and Change* (Chicago: University of Chicago Press, 1969), 147–78; Michael Hechter, *Internal Colonialism: The Celtic Fringe in British National Development, 1536–1966* (London: Routledge and Kegan Paul, 1975).

[5] Gourevitch, "The Reemergence of 'Peripheral Nationalisms': Some Comparative Speculations on the Spatial Distribution of Political Leadership and Economic Growth," *Comparative Studies in Society and History* 21 (July 1979), 303–22, at 319–21. See also Donald L. Horowitz, "Patterns of Ethnic Separatism," *Comparative Studies in Society and History* 23 (1981), 165–95.

Table 1
Major Union-Republic Legislation on Ethnic Relations
(September 1, 1985–December 31, 1989)

Republic	*Sovereignty*[a]	*Language*[b]	*Other Legislation*[e]
Estonian SSR	Constitutional amendment (11/16/88)[d]	Law (1/18/89)	Republican economic accountability (5/18/89)
			Voting residence requirement (8/8/89)[d]
			Deputy residence requirement (11/17/89)
Lithuanian SSR	Constitutional amendment (5/18/89)[d]	Law (11/18/88)	Republican economic independence (5/18/89)
		Presidium decree (1/26/89)	Law on citizenship (11/3/89)[d]
		Supreme Soviet decree (5/18/89)	Military service (9/29/89)[d]
Latvian SSR	Constitutional amendment (7/29/89)[d]	Supreme Soviet resolution (10/6/88)	In-migration restrictions (2/14/89)
		Law (5/5/89)	Voting residence requirement (7/29/89)[d]
Azerbaidjan SSR	Constitutional amendment (9/23/89)[d]	[Article 73][c]	
Georgian SSR	Constitutional amendment (11/19/89)	[Article 75][c]	
Armenian SSR		[Article 73][c]	Annexation of Nagorno-Karabakh (12/1/89)
Tadjik SSR		Law (7/22/89)	
Moldavian SSR		Law (9/1/89)	
Uzbek SSR		Constitutional amendment (10/21/89)	

Sources: *Izvestiia*, January 20, 1989, July 23, 1989, November 5, 1989, November 13, 1989, November 20, 1989; *Kommunist* [Erevan], December 3, 1989; *Pravda*, October 7, 1988, November 28, 1988, May 6, 1989, May 20, 1989, July 30, 1989, August 10, 1989, August 17, 1989, September 1, 1989, September 3, 1989, October 6, 1989; *Sovetskaia Estoniia*, November 19, 1988; *Sovetskaia Litva*, November 19, 1988, January 27, 1989, *Sovetskaia Latviia*, February 17, 1989

[a] Declarations of sovereignty and right of nullification. In parentheses. date of adoption.

[b] Declaration of state language. In parentheses: date of adoption.

[c] Language provision previously adopted with 1978 union-republic constitution.

[d] Portions subsequently declared unconstitutional by all-union Presidium of the Supreme Soviet

[e] In parentheses: date of adoption.

TABLE 2
MAJOR PEACEFUL DEMONSTRATIONS
(SEPTEMBER 1, 1985–AUGUST 31, 1989)

hnic Group	*Estimated Number of Demonstrations* Over 100,000 Participants	Over 10,000 Participants[a]
rmenians	25	30
zeris	9	19
thuanians	4	9
itvians	3	7
:orgians	2	4
tonians	2	4
oldavians	1	6
zbeks	1	1
:xclave" Russians	0	3
lorussians	0	1
crainians	0	0
izakhs	0	0
rgizes	0	0
idjiks	0	0
irkmen	0	0

JRCES: *New York Times; Radio Liberty Research Bulletin; Report on the USSR.*
Number includes those over 100,000.

ion Moscow for its members' positions. As this cadre was assigned a onopoly over the mobilizational resources within the ethnic commu- :y, it determined when the ethnic group would be mobilized to action. was a strategy that achieved interethnic peace not so much by remov- ; the root causes of ethnic grievances as by eliminating mobilizational portunities for independent ethnic protest.

It is therefore an ironic twist that after the transition to industrialism :se federal institutions and indigenous cadres became instruments of : new ethnic assertiveness. Institutions that were designed to expand oscow's control over ethnic groups (and that were generally thought the West to be moribund as federal guarantees of ethnic rights) have cen on a new life. Autonomous homelands provide essential resources · the collective mobilization of ethnic communities, and both federal ititutions and indigenous cadres shape ethnic agendas.

Central to what we have witnessed in the Soviet Union is an expand- ; and increasingly public politics of *ethnofederalism*. This is not a new enomenon initiated by the policies of Mikhail Gorbachev but the con- uation of a trend that began to unfold as early as the rule of Nikita

(a) Completed Higher Education, 1970
(Proportion of 10+ Age Group)

(b) Enrollment in Higher Education, 1970
(Proportion of 17-29 Age Group)

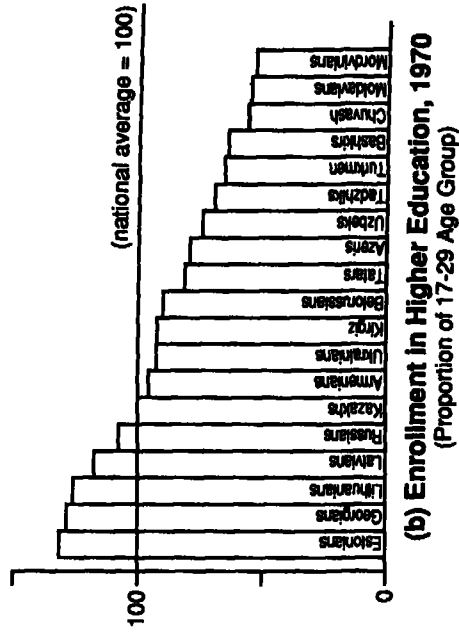
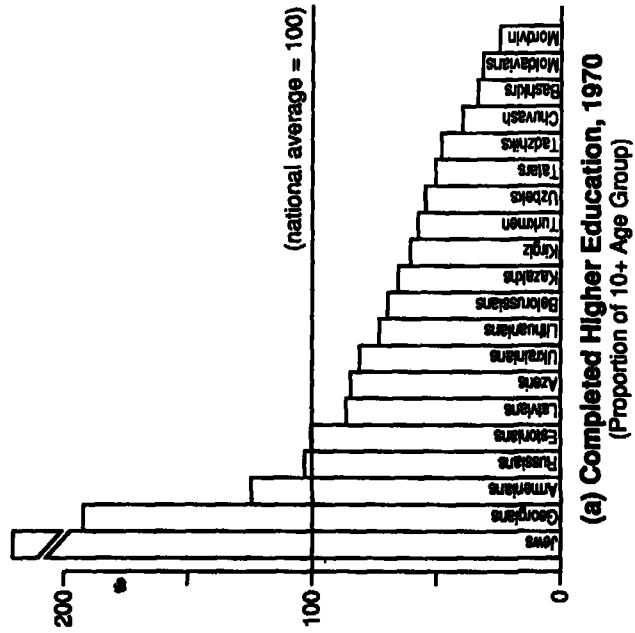
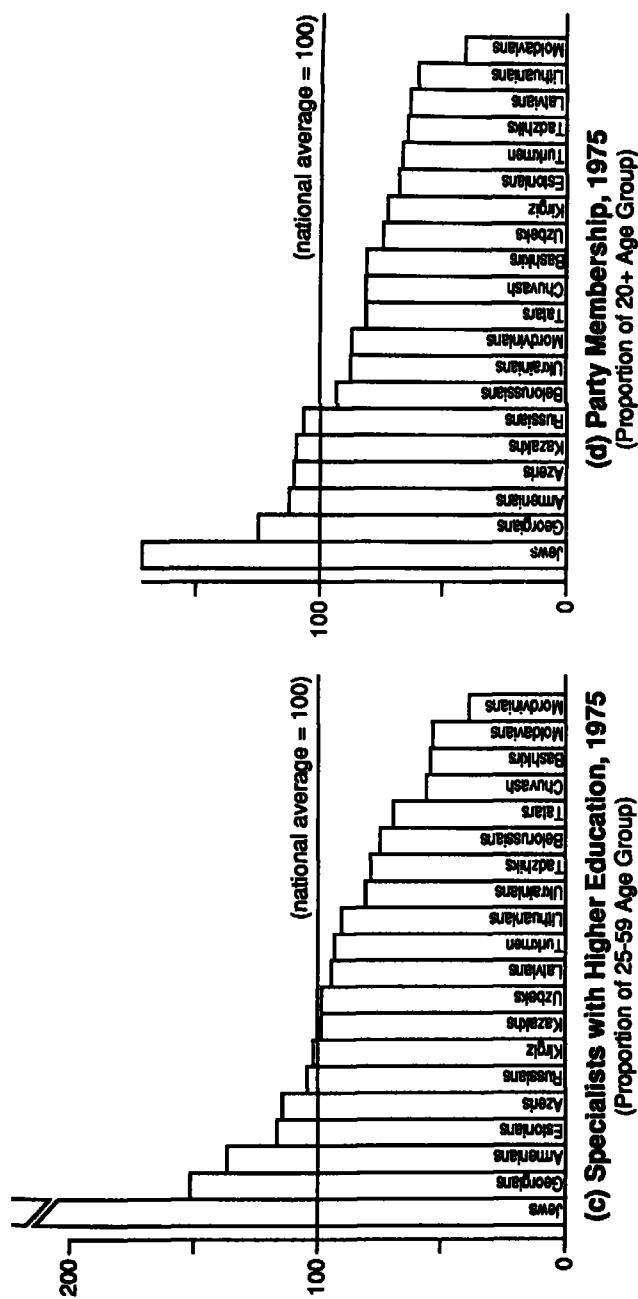



FIGURE 1

EDUCATIONAL, OCCUPATIONAL, AND POLITICAL ATTAINMENT BY ETHNIC GROUP[a]

SOURCES: Tsentral'noe Statisticheskoe Upravlenie, *Itogi vsesoiuznoi perepisi naseleniia 1970 goda* (Moscow: Statistika, 1972–73), 2:12–75, 4:360–82, 393–433; *Narodnoe obrazovanie, nauka, i kul'tura v SSSR* (Moscow: Statistika, 1971), 196; *Narodnoe obrazovanie, nauka, i kul'tura v SSSR* (Moscow: Statistika, 1977), 296; *Partiinaia zhizn'*, no. 10 (1976), 16.

[a] These figures show the most recent data available for ethnic groups with a population over one million.

Khrushchev. As examples below will show, demands for expanded autonomy, protests over language policy, pressures to reduce Russian migration, and intercommunal violence have surfaced in every decade since the mid-1950s. Gorbachev's policies are clearly not sufficient explanation for a pattern that predates Gorbachev and the introduction of his reforms.

The key to three questions raised by the present ethnic crisis lies within the Soviet developmental strategy that created those homelands and cadres. First, *origins*: Why have institutions that fostered interethnic peace during the transition to industrialization later become the vehicles of protest? Second, *incidence*: Why have the relatively advantaged ethnic groups been the most assertive, whereas ethnic groups near the lower end of most comparative measures of socioeconomic and political success have been relatively quiescent? And third, *agendas*: Why have the most important issues of contention between center and periphery focused to such a large degree upon the details of the Soviet developmental strategy and upon federalism in particular? As is argued in the conclusion, the answers to these questions point up the centrality of Soviet political institutions to the politicization of ethnicity.

The Soviet Developmental Strategy and Ethnicity

Political entrepreneurs play a critical role in the mobilization of protest, the politicization of ethnicity, and in many cases even the creation of ethnic identities.[6] In the European experience, for example, regional intellectuals who felt their aspirations to elite status frustrated by the status quo were often the pioneers of ethnic revival. Anthony Smith contends that these intellectuals sought to create a separate, ethnically distinct system of stratification within which the "professional and bureaucratic apparatus would naturally satisfy the career aspirations of a multitude of hitherto excluded diploma-holders."[7] These regionally oriented intellectuals were a necessary, although not sufficient, ingredient in the rise of peripheral nationalisms.

Political entrepreneurs in ethnic communities have available to them two mobilizational strategies: *primordial* and *instrumental*. The primor-

[6] William Bernard, "New Directions in Integration and Ethnicity," *International Migration Review* 5 (Winter 1971), 464–73; Michael Hechter, Debra Friedman, and Malka Appelbaum, "A Theory of Ethnic Collective Action," *International Migration Review* 16 (Summer 1982), 412–34; Phillip M. Rawkins, "An Approach to the Political Sociology of the Welsh Nationalist Movement," *Political Studies* 27 (September 1979), 440–57; Joseph Rothchild, *Ethnopolitics: A Conceptual Framework* (New York: Columbia University Press, 1981), 27.

[7] Smith, *The Ethnic Revival* (Cambridge: Cambridge University Press, 1981), 87, 126.

dial strategy focuses on ethnic revival—in Smith's words, "communal regeneration through self-discovery and self-realization."[8] The mobilization of the ethnic community for political action often centers on an assertion of the ethnic group's identity, usually in the context of issues of culture, identity, or belief and in reaction to threats to the identity from assimilative policies. The instrumental strategy focuses on the pursuit of social and economic interests. The mobilization of ethnicity, according to Rothchild, is "a highly conscious, political, and new mode of interest articulation."[9] The ethnic group itself, in the words of Nathan Glazer and Daniel Moynihan, is "defined in terms of interest, *as* an interest group."[10] Political entrepreneurs may also seek to mix these strategies.

The Soviet developmental strategy sought to control ethnopolitics by prohibiting all but sanctioned political entrepreneurs from mobilizing their communities and by deterring these entrepreneurs from pursuing any but the regime's instrumental strategies of plan fulfillment and social transformation. The Soviet strategy achieved this control through a threefold policy of (1) creating within each ethnic homeland an indigenous cadre assigned a monopoly over the mobilizational resources of the community, (2) constraining the behavior of this new ethnic cadre by creating an incentive structure that deterred the expression of unsanctioned, particularly primordial ethnic agendas, and (3) assigning the cadre the responsibility for creating an ethnically distinct stratification system within official institutions and for impeding the emergence of alternative ethnic entrepreneurs outside these institutions. Let us examine each in turn.

Creating an Ethnic Cadre

The Soviets have labeled the structural foundations of their "nationalities policies" socialist federalism and indigenization (*korenizatsiia*). The socialist federation, in the formulation of the *Great Soviet Encyclopedia*, differs radically from the bourgeois federation," for the former is "the state form for solving the *national question* ... [and] is based on the national-territorial principle."[11] Thus, at present fifty-three of the territorial administrations of the Soviet Union are based on designated ethnic homelands—fifteen as union republics, twenty as autonomous republics,

[8] Ibid., 105; James McKay, "An Exploratory Synthesis of Primordial and Mobilizationalist Approaches to Ethnic Phenomena," *Ethnic and Racial Studies* 5 (October 1982), 395–420, at 399.

[9] Rothchild (fn. 6), 30.

[10] Glazer and Moynihan, *Ethnicity: Theory and Experience* (Cambridge: Harvard University Press, 1975), 7.

[11] *Bol'shaia Sovetskaia Entsiklopediia* (Moscow: Sovetskaia entsiklopediia, 1977), 27:255.

eight as autonomous oblasts, and ten as autonomous *okrugs*.[12] Indigenization has sought to tie the minorities to the Soviet regime by drawing national cadres into the political and administrative posts of Party and state in these territories. In 1920 then People's Commissar of Nationality Affairs Joseph Stalin explained that to make Soviet power "near and dear" to the minorities would require

> that all Soviet organs in the border regions... should as far as possible be *recruited from the local people acquainted with the manner of life, habits, customs, and language of the native population;* [and] that all the best people from the local masses should be drawn into these institutions.[13]

Federalism and indigenization came at the expense of simple economic rationality and assimilation. The national-territorial principle has not always led to optimal administrative units. In some instances, particularly among less modernized groups, it perpetuated or strengthened ethnic differences that might otherwise have disappeared. In the case of the Ukraine, Alexander Motyl can ask, "Why ... did the Soviet state... discourage Little Russianization [i.e., assimilation] by pursuing *korenizatsiia*?"[14] The answer would appear to be the primacy the Soviet regime placed on checking the mobilizational sources of ethnopolitics and the critical role it assigned the ethnic cadres in this strategy.

These policies provided opportunities for nationalities representing over 93 percent of the non-Russian population to create ethnically distinct political elites within formally autonomous homelands. Grey Hodnett's extensive data show that by the early post-Stalinist period (1955–72), indigenization in eleven of the fourteen non-Russian republics led to proportionate overrepresentation of the titular nationality in Party and state leadership posts at the republic level. By the 1980s indigenization extended well beyond the most visible posts, such as each republic's Party first secretary, chairmen of its Presidium and Council of Ministers, first secretary of its Union of Writers, president of its Academy of Sciences, rectors of its principal universities, and chair of its council of trade unions. The data compiled by Ellen Jones and Fred W. Grupp show that it also reached such sensitive and less visible areas as internal security, including each republic's Ministry of Internal Affairs, Committee on

[12] A new (eleventh) autonomous *okrug* was created for the Even-Batagai district in Yakutia in October 1989; *Pravda*, October 30, 1989. Other changes are likely to follow.

[13] Stalin, "The Policy of the Soviet Government on the National Question in Russia," in *Works* (Moscow: Foreign Languages Publishing House, 1953), 4:370–71.

[14] Motyl, *Will the Non-Russians Rebel?* (Ithaca, N.Y.: Cornell University Press, 1987), 104. See also Grey Hodnett, "The Debate over Soviet Federalism," *Soviet Studies* 18 (April 1967), 458–81; Daniel C. Matuszewski, "Nationalities in the Soviet Future: Trends under Gorbachev," in Lawrence C. Lerner and Donald W. Treadgold, eds., *Gorbachev and the Soviet Future* (Boulder, Colo.: Westview Press, 1988), 95–96.

State Security, and the Party's Administrative Organs Department. It also touched lower levels of administration. In 1988, for example, in each union republic with oblasts the indigenous nationality held a greater proportion of oblast Party committee first secretaryships than its proportion of the republic's population.[15]

The indigenous cadre was given an institutionalized monopoly on the public expression of ethnic identity, that is, it defined the ethnic markers that distinguish the nationality. These markers were then central to communicating the socialist message in national cultural forms and propagandizing populations being brought into the modern sector. For many Soviet citizens undergoing social mobilization the first sustained contact with the great traditions of their own ethnic group was in the form of this national-Soviet hybrid. In the extreme, the markers identified by these elites defined new ethnic groups, such as the Tadjiks, that had not previously been communities with which elites and the masses had identified. Yet, as the recent political activism by Tadjiks attests, even these markers became the basis for the mobilization of the population in political action.[16]

More importantly, within each republic this cadre was assigned the role of gatekeeper, to determine when the ethnic group would be mobilized politically. Insofar as anyone within the homeland had access, this cadre monopolized the mobilizational resources essential to sustained, large-scale political action. The means of communications, particularly the indigenous-language press and broadcast media, were monopolized through the republican institutions controlled by this cadre.[17] Access to meeting places, such as auditoriums and public squares within the re-

[15] Hodnett, *Leadership in the Soviet National Republics* (Oakville, Ontario: Mosaic Press, 1978), 101–3, 377–78; Jones and Grupp, "Modernisation and Ethnic Equalisation in the USSR," *Soviet Studies* 36 (April 1984), 159–84, at 174; *Deputaty verkhovnogo soveta SSSR, odinadtsatyi sozyv* [Deputies of the Supreme Soviet of the USSR, eleventh convocation] (Moscow: Izvestiia, 1984); Gavin Helf, comp., *A Biographical Directory of Soviet Regional Party Leaders*, 2d ed. (Munich: Radio Free Europe/Radio Liberty, 1988). For long-term trends in indigenization within specific republics, see John A. Armstrong, *The Soviet Bureaucratic Elite: A Case Study of the Ukrainian Apparatus* (New York: Praeger, 1959), 15–17; Ronald Grigor Suny, *The Making of the Georgian Nation* (Bloomington: Indiana University Press, 1988), 209–318; Martha Brill Olcott, *The Kazakhs* (Stanford, Calif.: Hoover Institutions Press, 1987), 199–246. See also Steven L. Burg, "Russians, Natives, and Jews in the Soviet Scientific Elite: Cadre Competition in Central Asia," *Cahiers du Monde Russe et Soviétique* 20 (January–March 1979), 43–59; Nancy Lubin, "Assimilation and Retention of Ethnic Identity in Uzbekistan," *Asian Affairs* 12 (October 1981), 277–85, at 283; J. W. R. Parsons, "National Integration in Soviet Georgia," *Soviet Studies* 34 (October 1982), 547–69, at 554.

[16] Teresa Rakowska-Harmstone, *Russia and Nationalism in Central Asia* (Baltimore, Md.: Johns Hopkins University Press, 1970), 76; *Izvestiia*, July 14, 1989; *Kommunist Tadzhikistana*, June 28, 1989; *Pravda*, June 25, 1988.

[17] Alex Inkeles, *Public Opinion in Soviet Russia: A Study in Mass Persuasion* (Cambridge: Harvard University Press, 1950); Peter Kenez, *The Birth of the Propaganda State: Soviet Methods of Mass Mobilization, 1917–1929* (Cambridge: Cambridge University Press, 1985).

public, was at the discretion of this cadre. And public protests could avoid violent suppression only with the cadre's approval.

The Deterrence of Primordial Strategies

These cadres were encouraged to pursue the regime's instrumental strategies and deterred from primordial strategies by the offer of material rewards and status, which were tightly tied to the regime's goals.[18] Particularly after Stalin's purge of traditional native elites and the sovietization of indigenous institutions, the cadres enjoyed access to these rewards only by virtue of Soviet institutions. Their privileged positions would not be improved in alternative (even independent) institutions. Indeed, the collapse of Soviet power within their homeland would mean their own fall from power—or perhaps worse. Cadres thus had a strong incentive to resist the articulation of agendas that might be subversive of existing federal institutions.

Rewards were tightly tied to the norms and goals of the Soviet developmental strategy. Soviet federalism embedded these cadres within the all-union Soviet administrative hierarchy. Cadres could only succeed within an incentive system that defined individual and collective success in instrumental terms of quota fulfillment and socioeconomic growth. By integrating the cadres into Party and state hierarchies, Soviet federalism made the "normal" politics of competitive appeals for resources the norm among ethnic elites as well. Much of the politics between Moscow and the nationality-based territorial units came to involve the petitioning for funds from above; beginning with the last years under Khrushchev speeches by leaders of the republics appealing for funds and projects to benefit their people came to be an increasingly prominent feature of meetings of the Party Congress, the Central Committee, and the Supreme Soviet.[19]

Those who engaged in unsanctioned mobilizational strategies could be punished with total deprivation of these rewards by being removed from their positions of authority. The monopoly of official institutions meant that a purge at the very least threatened one's access to these rewards; cadres could not return to a prosperous private life. Purges of ethnic leaders charged with articulating particularistic, primordial agendas also deterred others from making such appeals. Since 1960 over a dozen first secretaries of union republics have been removed under cir-

[18] Seweryn Bialer, *Stalin's Successors* (Cambridge: Cambridge University Press, 1980), 216. See also Motyl (fn. 14), 104–5, 119–22.

[19] Donna Bahry, *Outside Moscow. Power, Politics, and Budgetary Policy in Soviet Republics* (New York: Columbia University Press, 1987), 1–5; Jerry F. Hough and Merle Fainsod, *How the Soviet Union Is Governed* (Cambridge: Harvard University Press), 510–17.

cumstances that suggest the cause was either their own endorsement of primordial agendas or their unwillingness to silence others who articulated such agendas. For example, after an official reception in 1962 the chairman of the Council of Ministers of the Kazakh Republic was apparently removed for making unguarded comments while intoxicated—comments that were just too nationalistic. In 1966 the first secretary of the Armenian Party was removed for failure to curb anti-Turkic protests on the fiftieth anniversary of the massacre of the Armenians. And in 1972 the Ukrainian and Georgian secretaries were both removed for "national narrow-mindedness" and overzealous promotion of local interests.[20]

Creating an Official Indigenous Elite

A major responsibility assigned this new cadre by Moscow's all-union authorities was to block the emergence within the ethnic community of counterelites that might challenge Soviet institutions. Within their autonomous homelands the cadres implemented policies that extended the institutional and personnel strategies of the center. These policies sought to (1) create a new, open indigenous elite of professionals and intelligentsia within official institutions, (2) tie professional and material rewards to membership in this elite, while denying these rewards to those outside the elite, and (3) limit access to the mobilizational resources of the community to these official institutions.

The ethnic cadres enacted affirmative action policies during a period of rapid economic growth and modernization in order to expand opportunities for mobility for those aspiring to positions within the professional strata and intelligentsia. Extension of the policy of indigenization opened career opportunities throughout the administrative apparatus of the homeland. Programs of collectivization and industrialization offered opportunities for mobility in management. And the creation of universities and academies of sciences in the republics dramatically expanded the number of professional positions reserved for the minorities.

These affirmative action policies in the institutional context of Soviet federalism elevated titular nationalities to privileged positions in higher education and professional employment within their homelands. For example, whereas Georgians constituted 67 percent of their republic's population in 1970 (and approximately the same proportion of the college-age cohort), they constituted 83 percent of the student body of the republic's institutions of higher education.[21] Similarly, although Mol-

[20] Robert Conquest, *Soviet Nationalities Policy in Practice* (New York: Praeger, 1967); Teresa Rakowska-Harmstone, "The Dialectics of Nationalism in the USSR," *Problems of Communism* 23 (May–June 1974), 1–22, at 13.

[21] Parsons (fn. 15), 558–59.

davians constituted under two-thirds of the total population of their republic in the mid-1980s, they were at least 80 percent of the student body in the law and business schools of Kishinev State University, the republic's leading educational institution. Commenting on the rapid upward mobility of the Uzbek population within their republic, Nancy Lubin contends that the Central Asians "tend to hire 'their own' first."[22]

It is probably true—as critics have charged—that these opportunities have been opened by crudely implemented quota systems. These have apparently lowered standards in higher education and employment and discriminated against "minority" ethnic groups living within the homelands of other ethnic groups. Mark Popovsky alleges that in the universities of Uzbekistan, for example,

> young men and women from primitive villages with scarcely any education . . . are given scholarships and free lodging, and are assured passing marks whether they study or not. The philosophy behind this strange proceeding is that all nations in the brotherly family of the USSR are equal, and that all of them therefore can and must have their own intelligentsia, their own doctors, engineers, writers, and scholars.[23]

And Soviet officials themselves have warned that favoritism in appointments toward the titular nationality of a republic often discriminates against other nationalities. In 1986, for example, the all-union Party leadership chastised Kirgiz leaders for favoring Kirgiz candidates among new recruits to the republic's Party organizations and for discriminating against other nationalities (notably Uzbeks) residing within the republic.[24]

Nonetheless, these policies, to expand the opportunity for mobility, represent a particularly astute accommodation with ethnicity. From 1950 to 1975, for example, among the fourteen titular nationalities of union republics (other than Russians) the annual growth in scientific workers with either a candidate of science or a doctor of science degree was 9.6 percent—a rate 54 percent higher than among Russians.[25] While demanding political loyalty to the Soviet regime, the mobility opportunities did not require denial of ethnic identities. Indeed, ethnicity became a condition for success, since the positions of status within homelands were often reserved for specific minorities. Soviet federalism offered minori-

[22] Lubin (fn. 15), 283.

[23] Mark Popovsky, *Manipulated Science* (Garden City, N.Y.: Doubleday, 1979), 118.

[24] John Soper, "Nationality Issues under Review in Kirgizia," *Radio Liberty Research Bulletin* RL 49/88 (January 29, 1988).

[25] USSR Tsentral'noe Statisticheskoe Upravlenie, *Narodnoe obrazovanie, nauka i kul'tura v SSSR* [Popular education, science and culture in the USSR] (Moscow: Statistika, 1977), 308–39.

ties the opportunity to realize their aspirations to create separate stratification systems—but within the Soviet Union.

To block the emergence of counterelites *within* this official elite, the cadres presided over a dense network of parallel institutions that controlled all aspects of professional life.[26] Research of significance to the ethnic group and its homeland was controlled by indigenous academies of sciences and universities. Creative professionals such as writers, artists, or architects who sought to disseminate their work under the cultural monopoly of the regime were required to join the official unions of the homeland for their respective professions. Members of the new scientific and creative elite were dependent upon the official institutions, and most could not hope to improve their lot in an alternative ethnic elite created outside and in opposition to these institutions. This was particularly true for those in positions of power such as leaders of professional unions, those with academic positions in fields heavily encumbered by ideology, and creative artists who depended upon the hegemony of socialist realism for their success. The purge threatened deprivation of both rewards and the means to practice one's profession.

To block the emergence of counterelites *outside* the new professional elite and intelligentsia, the cadre denied those outside official institutions access to mobilizational resources. The ethnic cadre prohibited independent association, severed unofficial lines of communication between the intelligentsia and the populace, and deprived incipient dissident movements of their leadership by "decapitating" them—that is, threatening, imprisoning, or executing exemplary figures. Those outside were limited to the inefficient means of illegal associations, samizdat, and underground dissemination. Muslims opposed to the official religious hierarchy in the mid-1980s, for example, still had to rely upon religious tracts—many of them handwritten—smuggled across the borders from Afghan resistance groups. And Muslims of the Sufi underground borrowed a technique from dissidents in the European parts of the Soviet Union: they distributed their religious texts by chain letters.[27] Soviet policies punished severely those who attempted to articulate primordial agendas outside the official institutions. For example, in an attack on what *Kommunist Ukrainy* labeled the "belching of debilitated national-

[26] See, for example, Dietrich A. Loeber, "Administration of Culture in Soviet Latvia," in Adolf Sprudz and Armin Rusis, eds., *Res Baltica* (Leiden: A. W. Sijthoff, 1968), 133–45; Nicholas P. Vakar, *Belorussia: The Making of a Nation* (Cambridge: Harvard University Press, 1956), 150–51.

[27] Timur Kocaoglu, "Muslim Chain Letters in Central Asia," *Radio Liberty Research Bulletin* RL 313/83 (August 18, 1983); Alexandre Bennigsen, "Mullahs, Mujahidin, and Soviet Muslims," *Problems of Communism* 33 (November–December 1984), 28–41, at 36–37.

ism," the KGB in August–September 1965 arrested over twenty intellectuals who had attempted to lead a Ukrainian cultural renaissance.[28] In order to crack down on nonofficial Muslim preaching, a 1982 decree of the Turkmen Supreme Soviet Presidium authorized sentences of up to two years imprisonment and corrective labor for the crime of "social vagabondage."[29] And in Armenia arrests in late 1988 and early 1989 sought to decapitate the protest movement by targeting its leaders (such as the Karabakh Committee).[30]

Consequences for Ethnic Political Action

As a consequence of the monopoly over mobilizational resources held by the official institutions, in most circumstances only instrumental political action behind the objectives of the cadre could muster the mobilizational requisites for sustained large-scale action. There were, of course, spontaneous incidents of primordial ethnic protest. In Tashkent in September 1969, for example, an Uzbek crowd assaulted Russian bystanders after a match between the local "Pakhtar" soccer team and visitors from the Russian Republic.[31] And many individual participants in officially sponsored political action harbored private primordial agendas. Nevertheless, aspiring counterelites were handicapped in their attempts to form and mobilize effective political action in support of primordial agendas; and when it occurred, primordial protest was more likely to be expressed in isolated, ineffective, small-scale events.

Even in the period of perestroika, in republics where these cadres exercise decisive control over mobilizational resources, they can determine whether protest will be on a sustained large scale or simply sporadic. This is illustrated by the differing fortunes of popular fronts in the Ukraine and the Baltic republics during 1988 and 1989. In the Ukraine, according to an American correspondent, "the hard line of the [union-republic] Communist leadership here is one reason the rise of Ukrainian self-consciousness has been slower than the surge of nationalism in the Baltic republics."[32] In the Baltic republics, by contrast, the cadre made

[28] Ludmilla Alexeyeva, *Soviet Dissent* (Middletown, Conn.: Wesleyan University Press, 1985), 31.

[29] Bennigsen (fn. 27), 40.

[30] *New York Times*, November 25, 1988, November 26, 1988, December 22, 1988, January 2, 1989. See also *Pravda*, January 26, 1990.

[31] Michael Rywkin, *Moscow's Muslim Challenge* (Armonk, N.Y.: M. E. Sharpe, 1982), 121; B. Brown, "Kazkhstan in 1987: The Year after Alma Ata," *Radio Liberty Research Bulletin* RL 5/88 (December 23, 1987), 2. See also Ronald Grigor Suny, *Armenia in the Twentieth Century* (Chico, Calif.: Scholars Press, 1983), 78–80.

[32] *New York Times*, March 9, 1989; see also June 20, 1988, August 24, 1988, August 25, 1988, October 10, 1988, November 30, 1988, December 2, 1988, December 7, 1988, February 5, 1989, April 9, 1989, April 14, 1989, April 22, 1989; Bohdan Nahaylo, "Baltic Echoes in the

available to the popular fronts the mobilizational resources of the Party and state. Thus, the founding of the Latvian Popular Front was attended by the republic's Party leadership, which applauded its call for autonomy. Statements by the national fronts were published by the official press. Their newspapers were printed by the official publishing houses. Their programs appeared on state-run television, including two full days of live coverage of the Sajudis Conference in 1988. Their meetings were sanctioned, including the assembly of the three Baltic popular fronts held at the Estonian Central Committee's House of Political Enlightenment. Their demonstrations received permits. As the chairman of the Sajudis Assembly, Vytautus Landsbergis, stressed in a 1989 interview with *Tygodnik Powszechny*, the ability of his movement to conduct its activities depended upon its close relationship with the republic's Party authorities.[33] The importance of these cadres is further illustrated by the republic's elections of 1990: the cadres in the Baltic republics permitted a choice among candidates with alternative ethnic agendas, whereas in other republics, such as Belorussia, cadres blocked this.[34]

The ethnic cadres have actually instigated many of the protests of the past few years. In the Baltic republics demonstrations have been orchestrated to support legislative initiatives on state languages and republican sovereignty. In November 1988 the Estonian Party leadership reportedly pressed members of the republic's Supreme Soviet for a unanimous vote and then orchestrated demonstrations on behalf of legislation to grant itself the power of nullification over all-union legislation. Similarly, in May 1989 Lithuanian Party leaders mobilized demonstrations in Vilnius to support legislation giving the republic power of nullification and then claiming exemption from the new all-union highway tax.[35] In Nagornyi Karabakh and Armenia demonstrations on behalf of Armenian annexation of the Karabakh began in early 1988 with the support of the local Party and state leadership—including formal legislative endorsement of their cause by local soviets. And apparently even much of the violence to force Azeris from Armenia (and Armenians from Azerbaidjan) that began in late November 1988 took place with the support of local Party and state officials.[36] Where leaders of a republic take decisive action to

Ukraine," *Report on the USSR* 1 (January 13, 1989), 18–20; idem, "Confrontation over Creation of Ukrainian 'Popular Front,'" *Report on the USSR* 1 (March 3, 1989), 13–17.

[33] *Tygodnik Powszechny*, February 5, 1989. See also *Izvestiia*, May 15, 1989; Saulius Girnius, "Unofficial Groups in the Baltic Republics and Access to the Mass Media," *Report on the USSR* 1 (May 5, 1989), 16–19.

[34] *New York Times*, March 5, 1990, March 6, 1990, March 31, 1990.

[35] Moscow TASS, November 20, 1988, reported in FBIS, *Daily Report: Soviet Union*, November 21, 1988, pp. 42–43; *New York Times*, November 17, 1988, November 27, 1988, December 8, 1988, May 19, 1989, May 25, 1989.

[36] *Kommunist* (Baku), October 14, 1988; *Izvestiia*, January 5, 1989.

block protests (as in the Ukraine until mid-1989) or to silence them (in Georgia after five days of protest in April 1989), future political actio is usually sporadic and small-scale.

The ability of different ethnic groups to mount large-scale, sustainea political action has thus been closely tied to the resources controlled by their cadres. The mobilizational opportunities are greatest among those cadres in control of a union republic or even an autonomous republic or oblast. At the other extreme, ethnic groups without an autonomous homeland or living outside it suffer a major handicap, though they may still find local mobilizing cadres and elites, as the Gagauz found among town and village authorities in southern Moldavia. "Exclave" Russians in Estonia and Moldavia have found these cadres among factory managers and subordinate Party officials in those republics.[17] Nationalities lacking even these resources suffer the greatest handicap in attempting to express their protest in any sort of sustained large-scale manner. Thus, the protests by Jews have been sporadic and small-scale: in 1988 and 1989, for example, the largest demonstrations by Jews drew only several hundred, and most demonstrations, only two to three dozen.[18]

Origins: The Rise of Ethnofederalism

Over the past three decades three changes have transformed Soviet developmental strategy into a source of ethnofederalism. First, with the trimming of the terror apparatus under Khrushchev and the policy of "respect for cadres" under Brezhnev, that is, the relaxation of Moscow's deterrent threat, these cadres had greater leeway in pressing their particularistic agendas.[19] Second, as the cadres built institutional and even popular support within their ethnic communities, their dependence upon the center declined. These new power bases enabled the cadres to take more assertive policy stands against Moscow. Third, these cadres encountered mounting difficulties in securing resources to continue the expansion of mobility opportunities within their homelands. In response, they often resorted to a strategy of mobilizing their elites behind legislative agendas

[17] *New York Times*, March 15, 1989; Vladimir Socor, "Politics of the Language Question Heating Up in Soviet Moldavia," *Report on the USSR* 1 (September 8, 1989), 33–36; *Pravda* July 29, 1989.

[18] See, for example, *Radio Liberty Research Bulletin* RL 43/88 (January 29, 1988), 5, RL 167/88 (April 15, 1988), 9, RL 177/88 (April 22, 1988), 10, RL 258/88 (June 17, 1988), 7; *Report on the USSR* 1 (May 19, 1989), 33.

[19] H. Gordon Skilling, "Group Conflict in Soviet Politics: Some Conclusions," in H. Gordon Skilling and Franklin Griffiths, eds., *Interest Groups in Soviet Politics* (Princeton: Princeton University Press, 1971), 399–405; Robert E. Blackwell, Jr., "Cadres Policy in the Brezhnev Era," *Problems of Communism* 28 (March–April 1979), 29–42.

and their populations in ethnic protest *in order to secure additional resources from Moscow and to maintain their hegemony within their* homelands. Ironically, the last two of these factors are natural consequences of the very means by which these cadres had previously controlled their ethnic communities on behalf of the center.

Policies of affirmative action permitted these cadres to build more secure political bases within their ethnic communities, for these polices created a loyal clientele. The creation of this clientele was also fostered by the post-Stalinist decentralization. As many administrative tasks were transferred from all-union to union-republic or republican ministries and as the cadres' discretion in personnel matters was expanded, the cadres' control of patronage opportunities was enlarged. Efforts to cement loyalties with the indigenous elite were aided—particularly during the Brezhnev years—by the lengthening term of office of these elites and by the reduction in the amount of rotation of personnel among homelands.[40] These power bases blunted the deterrent threat of the purge, for they made removal of a republic's first secretary more costly for Moscow. The removal of Dinmukhamed Kunaev brought two days of riots to Kazakhstan, for example. The ethnic constituencies developed by these ethnic elites became resources in showdowns with the all-union leadership over key policy choices.[41]

The cadres' motivation to mobilize their ethnic constituencies rose in recent decades as their monopolistic leadership within the ethnic community came under increasing threat. One threat has been the indigenous professional elite and intelligentsia itself. The very success of previous affirmative action policies created a large group with the skills to constitute themselves as independent political entrepreneurs. A second threat has come from the disparity between the increasing demands for rewards and mobility opportunities and the diminished capability to meet those demands. Given the large size of this elite and its already

[40] For example, the average term of union-republic first secretaries nearly doubled from 4.5 years on January 1, 1960, to 8.7 years ten years later, it then rose again to 12.0 years ten years after that. Grey Hodnett, *Leaders of the Soviet Republics, 1955–1972* (Canberra: Australian National University, 1973); Central Intelligence Agency, *Directory of Soviet Officials: Republic Organizations* (Washington, D.C.: Central Intelligence Agency, 1980); Hodnett (fn. 15), 63–65.

[41] Mark Beissinger, "Ethnicity, the Personnel Weapon, and Neo-Imperial Integration: Ukrainian and R.S.F.S.R. Provincial Party Officials Compared," *Studies in Comparative Communism* 21 (Spring 1988), 71–85; Patrick Cockburn, "Dateline USSR: Ethnic Tremors," *Foreign Policy* 74 (Spring 1989), 174–75; Suny (fn. 15), 301–5; Suny (fn. 31), 73–75; Joel Moses, "Regionalism in Soviet Politics: Continuity as a Source of Change, 1953–82," *Soviet Studies* 37 (April 1985), 184–211; Martha Brill Olcott, "Gorbachev's Nationalities Policy and Soviet Central Asia," in Rajan Menon and Daniel N. Nelson, eds., *Limits to Soviet Power* (Lexington, Mass.: Lexington Books, 1989), 77–81.

high levels of material rewards, it has become more difficult to provide the still larger material rewards necessary to ensure their continuing loyalty. And a third threat has resulted from the increasing difficulty encountered in further expanding the ethnic elite to accommodate new aspirants to this elite. The rapid growth of a large professional elite and intelligentsia has virtually saturated some ethnic communities with elite positions. Declining economic growth has compounded these problems over the past two decades. This threatened, first, the cadres' capacity to continue expanding the rewards of the elite and the mobility opportunities for new aspirants to elite positions and, second, their ability to offer the improved living standards necessary to keep the population tied to themselves rather than to alternative leaders. In short, the last three decades brought increasing threats to some of the very means by which the cadre had previously blocked the growth of counterelites and prevented the mobilization of the population behind primordial agendas. It is an irony of the Soviet developmental strategy that some of these threats are products of the very success of the means previously used to control ethnic assertiveness.

A crisis developed over the past decade as the number of ethnic groups experiencing this threat increased. Ethnic cadres were forced to intensify their pressure on Moscow to gain additional resources, and consequently, competition for the same scarce resources grew among ethnic communities. But declining growth rates left Moscow with even fewer resources to respond to rising demands. Faced with this crisis, ethnic cadres have found they must press Moscow still harder for investments and must devise strategies that are more clever yet to underscore the urgency of their agendas. For example, in pressing Moscow for expanded autonomy, Baltic Party leaders have turned to legislative showdowns and have mobilized popular demonstrations to support the position of the republics' elites.[42] Ethnic cadres have found enthusiastic allies in these strategies among their dependent elites and among aspirants to elite positions. Even aspiring independent political entrepreneurs pursuing primordial strategies often joined the political action mobilized by the cadres, viewing it as the best or only vehicle to press their own agendas in public. The cadres often have a strong incentive to make common cause with members of potential counterelites such as popular fronts in order to increase the pressure on Moscow. Thus, the late 1980s witnessed both

[41] See Bahry (fn. 19), 2–3, 25–31, 77–85; Steven L. Burg, "Muslim Cadres and Soviet Political Development: Reflections from a Comparative Perspective," *World Politics* 37 (October 1984), 24–47, at 33, 36.

dominant elites and potential counterelites making common cause to mobilize the ethnic community.

Incidence: Flash Points of Ethnofederalism

The pattern of ethnofederalism in the Soviet Union contains yet a further ironic twist: the incentive to mobilize their ethnic constituency is greatest for those ethnic cadres that have previously been most successful at the Soviet developmental strategy—notably in the Caucasus and the Baltic republics. The pressure of potential counterelites, the difficulties of further expanding elite positions and material rewards, and so the threat to their positions are greatest for those cadres that were previously most successful at engineering affirmative action and creating an indigenous elite.

This pressure on the cadres of the more developed ethnic communities has been piqued by the redistributive consequences of these affirmative action policies. With declining growth rates, the redistributive consequences of these policies have been transformed: policies that had once involved transfers between titular nationalities and minorities within ethnic homelands came to involve instead transfers between the principal titular nationalities of union republics. These transfers are now adversely affecting the more modernized ethnic communities

In building an indigenous cadre and intelligentsia within each ethnic group, the Soviet developmental strategy had a powerful leveling impact on ethnic groups. The growth of mobility opportunities has been highest among the nationalities with the lowest levels of socioeconomic attainment.[43] Thus, as Figure 2 shows, in the post-Stalinist years differences among nationalities in levels of elite occupational status (measured by per capita employment as specialists with higher education) narrowed.[44]

This redistribution has in part been a consequence of Moscow's allocation of resources among republics. For example, the Unified State Budget, which includes the budgets for each union republic, transfers funds from more developed to less developed republics. In the 1989 bud-

[43] USSR (fn. 25), 308–9.

[44] Ellen Jones and Fred W. Grupp, "Measuring Nationality Trends in the Soviet Union: A Research Note," *Slavic Review* 41 (Spring 1982), 112–22. This is not to say that equalization has brought equality or status reversal, see Rakowska-Harmstone (fn. 20), 12; Peter R. Zwick, "Soviet Nationality Policy: Social, Economic, and Political Aspects," in Gordon B. Smith, ed., *Public Policy and Administration in the Soviet Union* (New York: Praeger, 1980), 159. Moreover, as Figure 2 shows, the slowdown in economic growth has slowed (but not stopped) this leveling process.

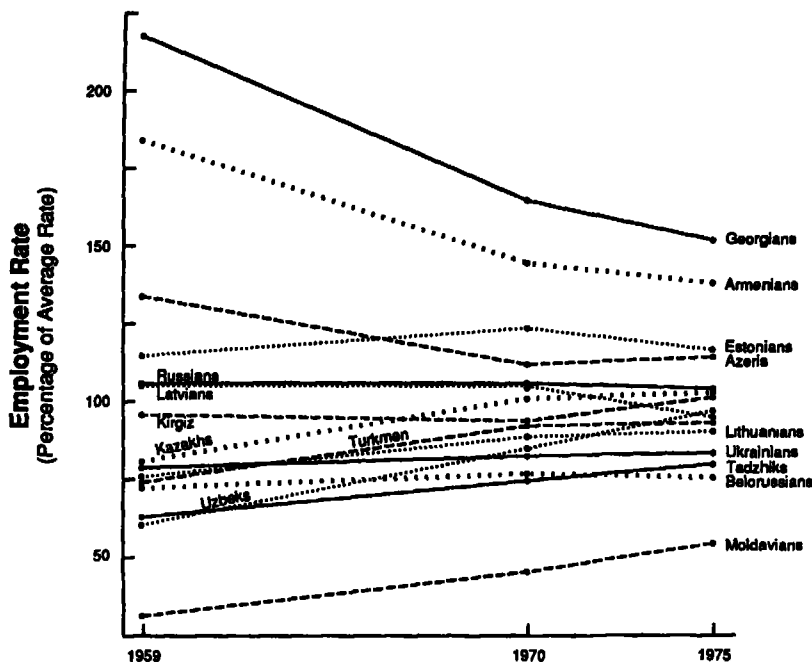



FIGURE 2
EQUALIZATION AMONG ETHNIC GROUPS, 1959–75:[a]
EMPLOYMENT OF SPECIALISTS WITH HIGHER EDUCATION PER 1,000 ADULTS
(AVERAGE FOR 15 NATIONALITIES = 100)

SOURCE: Calculated from data in Ellen Jones and Fred W. Grupp, "Modernization and Ethnic Equalization in the USSR," *Soviet Studies* 36 (April 1984), 159–84.
[a] This figure tracks elite employment in the Soviet economy in those years for which data are available for all titular nationalities of union republics.

get only the five Central Asian republics were permitted to retain 100 percent of both the turnover and income taxes collected within their borders; they were to receive, in addition, subsidies ranging from 321 million rubles for the Tadjik Republic to 2.7 billion rubles for the Kazakh Republic. Conversely, the Latvian Republic was to retain the lowest proportion of its turnover tax (56.8 percent), and both the Armenian Republic (with 76.7 percent) and the Estonian Republic (with 79.4 percent) were to retain only slightly over three-quarters.[45] Although these official statistics appear to overstate the extent of this phenomenon, less developed republics have received higher rates of investment than their level

[45] *Pravda*, October 29, 1988.

of economic development would predict. And per capita expenditures on health and educational programs have been relatively equal among republics even though revenues have been far lower in the less developed republics.[46]

It is, indeed, inherent to a policy of promoting the growth of indigenous elites and engineering equality under circumstances of tighter constraints on resources that some previously advantaged groups will suffer relative stagnation in their life chances. One of the most visible examples of this restriction on life chances for a previously advantaged group has been the imposition of quotas on Jewish admissions to Soviet universities. Although in 1970 Jews represented 2.3 percent of the students in institutions of higher education (a proportion that is above their official proportion of the population—0.89 percent in 1970), this, nonetheless, represents a significant decline since 1935, when the figure stood at 13.3 percent of these students. Against the general trend toward higher rates of educational attainment in the Soviet population, the proportion of Soviet Jews aged eighteen to twenty-three who attend college full-time remained constant at 30 percent or at most increased only marginally to 36 percent between 1935 and 1965. Thus, many have been blocked in their aspirations for higher education and the elite employment it would make possible.[47] And as economic growth has slowed, more ethnic groups have felt this pinch in life chances. In Central Asia, according to Lubin, "Russians are beginning to sense they are being denied access to jobs for which they are equally if not more qualified than their Asian counterparts."[48] And Popovsky complains that in Uzbekistan "it is almost impossible for non-Uzbeks with a higher education to get jobs."[49]

In the geographically segmented multiethnic society of Soviet federalism the redistributive consequences of these policies remained a less

[46] Martin Spechler, "Regional Development in the USSR, 1958–1978," in *Soviet Economy in a Time of Change*, U.S. Congress Joint Economic Committee (Washington, D.C.: Government Printing Office, 1979), 145. See also Donna Bahry and Carol Nechemias, "Half Full or Half Empty? The Debate over Soviet Regional Inequality," *Slavic Review* 40 (Fall 1981), 96; Elizabeth Clayton, "Regional Consumption Expenditure in the Soviet Union," *ACES Bulletin* 17 (Winter 1975), 35–43; James W. Gillula, "The Economic Interdependence of Soviet Republics," in *Soviet Economy in a Time of Change*, 629; Gertrude Schroeder, "Soviet Regional Policies in Perspective," in *The USSR in the 1980's* (Brussels: NATO Directorate of Economic Affairs, 1978), 131; Brian Silver, "Levels of Sociocultural Development among Soviet Nationalities: A Partial Test of the Equalization Hypothesis," *American Political Science Review* 68 (December 1974), 1637.

[47] William Korey, "The Legal Position of Soviet Jewry: A Historical Enquiry," in Lionel Kochan, ed., *The Jews in Soviet Russia since 1917* (London: Oxford University Press, 1972), 94–95; Alec Nove and J. A. Newth, "The Jewish Population: Demographic Trends and Occupational Patterns," in Kochan, 147.

[48] Lubin (fn. 15), 228, 283–84.

[49] Popovsky (fn. 23), 138.

contentious issue in federal politics as long as economic growth permitted the continued expansion of mobility opportunities for all titular nationalities within their homelands. With their high growth rates the titular nationalities of union republics could escape the consequence of interrepublic transfers by shifting the brunt of their impact to their minority populations—that is, by discriminating against them. Under these circumstances the minorities within the homelands of other nationalities (including the "exclave" minorities, such as Jews or Russians) were the greatest losers from interrepublic redistribution. The economic slowdown, however, made it more difficult for the titular nationalities to escape the effects of interrepublic transfers, which made redistribution a highly contentious issue among the cadres. The effect of the Soviet developmental strategy in a period of tight constraints on resources has been to limit the growth of mobility opportunities in the more developed ethnic communities in order to permit continued expansion in those that are less developed.

As a consequence, cadres of ethnic communities with higher levels of socioeconomic attainment (particularly in the Baltic and Caucasus) have led the way in pressing their ethnic legislative agendas, while cadres in less advanced communities (notably in Central Asia) have been less inclined to do so (see Table 1). As continuing beneficiaries of the developmental strategy enforced by the center, the latter cadres have often been harsh critics of the decentralization proposed by their peers in the more developed communities; for example, at the meeting of the all-union Supreme Soviet Presidium to veto Estonia's act of nullification, the chairmen of the Presidia of Uzbekistan (Khabibullaev) and Tadjikistan (Pallaev) voiced strong criticism of the Estonian move and support for centrist policies.[50] Also as a consequence, protest has been more common in the more developed communities. The rank-order correlation between number of demonstrations (Table 2) and levels of educational attainment (Figure 1a) is 0.67.[51]

Agendas: The Federal Politics of Resources and Life Chances

The policy concerns of the ethnic cadres shape the public agendas of ethnofederalism by controlling the way dominant themes will be framed in legislation and by determining which issues are to be supported by sustained large-scale political pressure. Cadres must define the agendas

[50] Moscow Television, November 26, 1988, reported in FBIS, *Daily Report: Soviet Union*, November 28, 1988, p. 50. See also James Critchlow, "How Solid is Uzbekistan's Support for Moscow?" *Report on the USSR* 1 (February 10, 1989), 7.

[51] See also Philip G. Roeder, "Electoral Avoidance in the Soviet Union," *Soviet Studies* 41 (July 1989), 478–80.

of ethnopolitics so as to permit as many as possible in their ethnic communities to join the "official" ethnic bandwagon. Often this requires redefining the dominant popular concerns, particularly where they are primordial. The cadres must deny these constituents and issues to potential counterelites, even as they use them to put pressure on Moscow. A test of the success of the entrepreneurship of the cadres is the extent to which they can insert the most important material and symbolic concerns of their constituents into the public agendas in ways that protect or promote their own power base. The dominant themes on the cadres' agendas reflect the threats that confront them within their own communities: these themes concern resources and life chances. Specifically, this politics of mobility opportunities is expressed in a number of recurring policy issues that go to the very foundations of the Soviet developmental strategy that created and sustain these cadres: federalism, indigenization, language, economic development, and migration.

Federalism

The most volatile issue of federalism has been the balance of power between center and periphery, since at stake are the resources controlled by the cadres and the mobility opportunities within their communities. Party leaders have used this issue to co-opt primordial concerns for independence to instrumental demands for expanded republican autonomy within Soviet federalism.[52] Expanded autonomy is one way for ethnic cadres to enlarge the resources within their control. Autonomy increases their discretion in the allocation of positions of power within the republic and in the administration of educational and occupational policies. And for cadres within the more developed republics that have felt most severely pinched by affirmative action, autonomy is a way to retain resources at home.

The cadres in the more developed union republics have been particularly quick to raise the banners of autonomy and sovereignty in order to blunt the redistributive consequences of all-union policies. Thus, for example, in 1989 Lithuanian leaders sought exemption from the all-union highway tax since the tax was levied on vehicles (which are more common in the relatively wealthy Baltic region) but spent disproportionately to build and repair roads in the less developed republics.[53] In all three Baltic republics calls for "regional economic accountability," "territorial

[52] *New York Times*, September 24, 1989; see also March 25, 1988; *Pravda*, November 2, 1988; *Sovetskaia Estoniia*, November 29, 1988.

[53] Moscow TASS, November 20, 1988, reported in FBIS, *Daily Report: Soviet Union*, November 21, 1988, pp. 42–43; *New York Times*, November 17, 1988, November 27, 1988, December 8, 1988, May 19, 1989, May 25, 1989.

cost accounting," and "self-financing" have supported the attempts the part of republican cadres to wrest control of industries from the ce tralized ministries. These plans envision the isolation of their marke from the larger economy of scarcity by such mechanisms as export ba riers to other union republics, identity cards to limit purchases by visitoi from other republics, prices determined at the republic level to improv the terms of exchange among republics, and even separate currencies foi some republics.[54]

For ethnic cadres of "minorities" subordinate to the union republic of another nationality, the issue of autonomy has often taken the form of demands over the "status" of their homelands within the federal hierarchy—that is, whether a homeland is a union republic, an autonomous republic, an oblast, or an *okrug*. The control of resources by the cadres of a nationality increases with this hierarchy. Thus, in a 1988 report on the mounting pressure and public rallies on behalf of the elevation of the Tatar Autonomous Republic to union-republic status, the writer Rafail Mustafin asked in the pages of *Pravda*:

> After all, what prompted the proposals for the creation of a new union republic? The existing inequality in social, political, and economic rights between union and autonomous formations. Tataria surpasses some union republics in both population and industrial potential. Yet, we don't have a feature-film studio of our own, we have only one publishing house (which is not very big), and there is an acute shortage of paper for publishing books, newspapers, and magazines.[55]

With similar objectives the Party leadership in Abkhazia has sought elevation of their autonomous republic to union-republic status, leaders of Moldavia's Gagauz minority have demanded creation of their own autonomous republic, and Polish leaders in the Shalchinin district of Lithuania have asked for autonomous status. In the last case, according to *Izvestiia*, "perhaps the sorest point was the impossibility for Poles to obtain higher education in their native language" within Lithuania.[56]

Taking a different tack, the cadres of other "subordinate" minorities

[54] *Izvestiia*, September 11, 1988, August 3, 1989, August 15, 1989; *Sovetskaia Industriia*, March 2, 1989; *Sovetskaia Litva*, October 7, 1988. See also *Izvestiia*, May 19, 1989; *Sovetskaia Litva*, May 19, 1989; *Pravda*, July 30, 1989; Dzintra Bungs, "A Comparison of the Baltic Declarations of Sovereignty," *Report on the USSR* 1 (September 15, 1989), 13–16; Kestutis Girnius, "The Lithuanian Communist Party and Calls for Sovereignty," *Report on the USSR* 1 (February 17, 1989), 18–20.

[55] *Pravda*, January 25, 1989.

[56] *Izvestiia*, September 7, 1989, November 14, 1989; *Literaturnaia Gazeta*, March 9, 1988. See also *New York Times*, April 8, 1989, April 9, 1989, April 16, 1989, August 7, 1989; and the interview with G. A. Pogosyan of the Nagornyi Karabakh, ArmenPress International Service, August 16, 1988, reported in FBIS, *Daily Report: Soviet Union*, August 23, 1988, p. 43.

have mobilized their populations to demand changes in the lines of authority among homelands. A nationality subordinate to the republic of a different titular nationality may find its resources and life chances limited. In Nagornyi Karabakh, Armenians protested that even though their autonomous oblast had the second-highest industrial production per capita within the republic, the leadership of the Azerbaidjan Republic allocated it lower than average per capita investments. G. A. Pogosyan, the first secretary of the autonomous oblast, is reported to have told the all-union Supreme Soviet Presidium that as a consequence, "even today, the Nagorno-Karabakh Autonomous Oblast does not have its own flour mill, nor its own concentrate-feed plant, nor its own reinforced-concrete structures plant, nor its own housing-construction combine." These investments lagged particularly in the social sphere. The Nagorno-Karabakh Armenians complained that inadequate investment in institutions of higher education within the oblast and Baku's opposition to letting Armenians attend institutions in the Armenian Republic meant that the Armenians within the oblast were denied access to higher education and career advancement.[57] Reassigning the Nagorno-Karabakh Autonomous Oblast to the Armenian Republic would put an end to the discrimination by Azeris and also open up union-republic mobility opportunities to its Armenian majority.[58] (Similar complaints and demands for removal from the jurisdiction of the Georgian Republic have come from the leaders of the Adzhar Autonomous Republic and the South Osetian Autonomous Oblast.)[59]

INDIGENIZATION

Cadres of titular nationalities have also mobilized political action in order to preserve the results of indigenization. Cadres have found that this issue harnesses both the instrumental interests of those in their ethnic community who aspire to elite positions and the primordial sentiments of those who see ethnic similarity as a requisite of legitimate authority. At stake for union-republic cadres are the very means by which they

[57] *Pravda*, July 20, 1988.

[58] On December 1, 1989, the Supreme Soviet of the Armenian Republic and the National Council of Nagornyi Karabakh jointly passed a resolution annexing the autonomous oblast to Armenia; *Kommunist* [Erevan], December 3, 1989.

[59] *Bakinskii Rabochii*, March 11, 1988; *Izvestiia*, March 25, 1988; *Krasnaia Zvezda*, February 28, 1989; *Moskovskie Novosti*, March 20, 1988; Roman Solchanyk and Ann Sheehy, "Kapitonov on Nationality Problems in Georgia," *Radio Liberty Research Bulletin* RL 125/78 (June 1, 1978), 1–5; Elizabeth Fuller, "How Serious are Inter-Nationality Tensions in Georgia?" *Radio Liberty Research Bulletin* RL 444/83 (November 25, 1983), 1-9; idem, "Abkhaz-Georgian Relations Remain Strained," *Report on the USSR* 1 (March 10, 1989), 25–27; Rasma Karklins, "Ethnic Politics and Access to Higher Education: The Soviet Case," *Comparative Politics* 16 (April 1984), 277–94, at 278.

have built their supporting clientele and popular constituencies. In Kazakhstan the removal of First Secretary Kunaev and the purge of the republic party apparatus in 1986 threatened the clientele network drawn disproportionately from the first secretary's own tribal group; it threatened the career prospects of the indigenous nationals who had tied their careers to the local leadership; and it brought a wave of violent protest.[60] *More recently, Georgian leaders have mobilized protesters to challenge Moscow's intrusion in the republic's personnel policy—particularly the efforts to challenge discrimination by the Georgians against minorities.*

"Minority" cadres within the national territories of others have mobilized their constituencies for the opposite end—to criticize abuses of indigenization by the titular nationalities and to call for Moscow's intervention against the practice. The creation of an ethnic elite and intelligentsia by the titular nationality often denies minorities within its national territory comparable opportunities. In a recent roundtable held under the auspices of the Research Council on Nationality Problems of the Academy of Science, E. V. Tadevosian complained that indigenization in some national republics has "often led to an artificial overrepresentation of the indigenous nationalities at the expense of other nationalities residing in those areas in the state organs, the administrative apparatus, the students and faculties of higher educational institutions, etc."[61] Gagauz in Moldavia and Nagorno-Karabakh Armenians in Azerbaidjan have demonstrated to protest discrimination in political appointments and career opportunities. The chairman of the presidium of the Abkhaz Supreme Soviet claims that such grievances led to violent disturbances in his republic not only in 1989, but also in 1957, 1967, and 1978.[62]

LANGUAGE

To preserve the foundation on which ethnic affirmative action was built, cadres have mobilized political action to raise or preserve the status of the language of their respective communities. Mobility opportunities are inextricably tied up with the status of indigenous languages. At stake for the constituents is their instrumental interest in privileged access to economic, social, and political power. For example, the language used in

[60] Brown (fn. 31), 1–4.

[61] I. D. Koval'chenko et al., "Natsional'nye protsessy v SSSR—itogi, tendentsii, problemy" [National processes in the USSR: Results, tendencies, problems], *Istoriia SSSR* (November–December 1987), 5–120, at 63, 73, 74, 79–80. See also Erwin H. Epstein, "Ideological Factors in Soviet Educational Policy towards Jews," *Education and Urban Society* 10 (February 1978), 227–28; Rasma Karklins, *Ethnic Relations in the USSR* (Boston: Allen and Unwin, 1986), 142, 146, 219.

[62] V. Kobakhia, Pis'mo [Letter], *Argumenty i fakty* (October 7–13, 1989), 8.

universities influences the admissions as well as later career opportunities of different ethnic groups within a republic. But language is also one of the most important primordial markers *defining an ethnic community.*

Through legislation that their languages supplant Russian as the language of communication within their republics, ethnic leaders in the union republics have sought to strengthen indigenous control over political and economic institutions. An article declaring the language of the titular nationality to be the state language of the republic was included in the 1978 constitutions of Armenia, Azerbaidjan, and Georgia. On April 14, 1978, an estimated five thousand Georgian demonstrators took to the streets to protest Moscow's attempt to amend this article in their republic's constitution. Indigenous cultural, educational, and scientific elites were particularly eager to maintain Georgian as the language of discourse in their professions. Georgian students joined the protests, as Ronald Suny notes, since the language clause ensured that "higher education in Georgia had become the prerogative of Georgians, and other nationalities found it difficult to enter schools of higher learning."[63] In 1988 and 1989 six other republics adopted legislation establishing a state language.

Minority cadres within national territories of others have mobilized political action to protest exactly this form of preference for the language of the titular nationality. At stake for the leaders of these minority communities is often the control of mobility opportunities, the future of the constituency on which they have built their power, or even their own positions. Protests against new language laws that would place them at a disadvantage have come from the Abkhazians in Georgia, "exclave" Russians and Poles in Lithuania, and both the Turkic Gagauz and the "exclave" Russians in Moldavia. Non-Estonian cadres in Estonia have called strikes to protest a law requiring that those who do not learn the language of the republic within four years be dismissed from their jobs: on March 14, 1989, as many as sixty thousand Russians and other "minorities" (according to *Izvestiia*'s count) took to the streets of Tallinn to protest "creeping counterrevolution endangering socialism in Estonia." Russian Party officials and factory managers within the republic reportedly mobilized their constituencies because they feared that the law of January 18, 1989, would cost them their positions.[64]

[63] Suny, "Geogia and Soviet Nationality Policy," in Stephen F. Cohen, Alexander Rabinowitz, and Robert Sharlet, eds., *The Soviet Union since Stalin* (Bloomington: Indiana University Press, 1980), 213, 219. See also *Izvestiia*, February 25, 1989; Charles E. Ziegler, "Nationalism, Religion and Equality among Ethnic Minorities: Some Observations on the Soviet Case," *Journal of Ethnic Studies* 13 (Summer 1985), 9–32, at 27.

[64] *Izvestiia*, March 17, 1989. See also *Izvestiia*, March 10, 1989, August 31, 1989; *Pravda*,

Development and Migration

To preserve the ethnically exclusive institutions on which their power rests, ethnic cadres have mobilized political action to oppose all-union economic development and migration policies. Although two different issues, they have become closely connected in public discussions since rapid industrialization has become the principal magnet for workers of other nationalities. For the population this threatens dissolution of the ethnic community, heightened competition for the leading political and economic positions in the republic, and loss of their affirmative action advantage in life chances. For the cadres this migration threatens the ethnic community on which their political power is based. It increases the pressure on the cadres to incorporate these other ethnic groups into the elite of the republic—pressure, that is, to dilute the ethnic homogeneity of the elites on which the cadres base their power. Failure of a titular nationality to maintain its numerical predominance within a republic might lead to the replacement of the ethnically exclusive cadre by one of more diverse ethnic composition; it might bring demotion to the status of an autonomous republic (as happened to the Karelo-Finnish Republic) or even outright dissolution of its autonomous homeland. Thus, in the early 1960s Georgian and Latvian officials opposed Khrushchev's plans for further expansion of heavy industry in the republics precisely because they feared it would bring still more Russian workers.[65] Armenian officials reportedly excluded Russian workers from permanent housing during the construction of the Erevan subway in order to prevent them from remaining. In the late 1980s protesters in the Baltic republics and Armenia demanded that their republics be given control of industrial and agricultural policy in order to slow the influx of outsiders. And Estonian cadres have attempted to blunt the effect of migration on their ethnic constituencies in future elections: legislation of the Estonian Republic disenfranchised those not resident in a district for two years (or in the republic for five years) and barred from office those not resident in a district for five years (or in the republic for ten years).[66]

August 17, 1989; *Sovetskaia Litva*, February 14, 1989; Karklins (fn. 59), 290–91; Roman Solchanyk, "Russian Language and Soviet Politics," *Soviet Studies* 34 (January 1982), 23–42; *New York Times*, June 21, 1988, June 22, 1988, July 23, 1988, September 2, 1988, January 19, 1989, July 27, 1989, July 28, 1989, August 7, 1989.

[65] Suny (fn. 63), 213; Juris Dreifelds, "Latvian National Demands and Group Consciousness since 1959," in George Simmonds, ed., *Nationalism in the USSR and Eastern Europe in the Era of Brezhnev and Kosygin* (Detroit: University of Detroit Press, 1977), 136–56; Jaan Pennar, "Nationalism in the Soviet Baltics," in Erich Goldhagen, ed., *Ethnic Minorities in the Soviet Union* (New York: Praeger, 1968), 206. See also *Pravda*, March 1, 1989.

[66] *Pravda*, July 29, 1989, August 10, 1989. See also *New York Times*, June 21, 1988, June 22,

THE RISKS OF ASSERTIVE FEDERALISM UNDER PERESTROIKA

With perestroika the homeland cadres press their agendas of ethnofederalism in a more complex environment. Within their homelands they are increasingly called upon to control and balance three very different forms of political action that draw together different sets of actors behind three often diverging agendas. Alongside the assertive federalism of the cadres is a second arena—the organizing activity and popular demonstrations of the popular fronts. Drawing particularly upon students, the intelligentsia, and the professional elite, the programs of the fronts often give expression to many of the primordial concerns of these ethnic communities. In a third arena communal violence draws heavily from the unemployed and from displaced refugees. Intercommunal violence has pitted Armenians against Azeris, Georgians against Abkhazians, and Georgians against Osetians in the Caucasus; young Uzbeks have attacked Meskhetis in the Fergana region, and Kazakh youths have attacked immigrants in the Novyi Uzen region of Central Asia (see Table 3).[67] The ethnic cadres in a number of republics have sought to use the political pressure of the second set of actors to support its showdowns with Moscow. Even without endorsing all the particulars of the fronts' agendas, the cadres have found that the fronts strengthen their own

TABLE 3
DEATHS FROM COMMUNAL VIOLENCE
(SEPTEMBER 1, 1985–AUGUST 31, 1989)

Ethnic Groups	*Estimated Total Dead*
Armenian-Azeri	>110
Uzbek-Meskheti	>100
Georgian-Abkhazian	14
Kazakh–Non-Kazakh	5
Tadjik-Meskheti	2

SOURCE: *New York Times*, June 21, 1989, June 25, 1989, June 26, 1989, July 17, 1989, September 5, 1989, September 17, 1989.

1988, June 23, 1988. After this was nullified by the All-Union Supreme Soviet Presidium, Estonia adopted new legislation requiring that candidates for republic positions must have been residents of Estonia for at least ten years; *Pravda*, November 18, 1989.

[67] *Izvestiia*, June 9, 1989, June 10, 1989, June 20, 1989, June 23, 1989, December 1, 1989; *Pravda*, June 5, 1989, June 6, 1989, June 7, 1989, June 10, 1989, June 12, 1989, June 20, 1989, June 25, 1989, January 17, 1990; Annette Bohr, "Violence Erupts in Uzbekistan," *Report on the USSR* 1 (June 16, 1989), 23–25; Ann Sheehy, "Interethnic Disturbances in Western Kazakhstan," *Report on the USSR* 1 (July 7, 1989), 11–14.

hands in negotiations with Moscow by permitting them to argue convincingly that their hands are tied at home. In some instances cadres have even encouraged the third form of political action in order to solidify their hold within the homeland and to press Moscow. In Azerbaidjan, for example, local Party leaders reportedly encouraged the growth of the popular front and helped found the more militant and violence-prone National Defense Committee.[68]

The ethnic cadres run immense risks, for these strategies may undermine their positions or even unleash forces they cannot control. Moscow, for example, could respond with coercion, as it did in 1988 and 1989, when the all-union leadership removed over twenty-five hundred officials (including the union-republic first secretaries) accused of abetting the intercommunal violence in Armenia and Azerbaidjan.[69] Cadres with strong ethnic constituencies may gamble that Moscow will be unable to remove them or will hesitate to pay the costs associated with such actions. The extensive power base of Vladimir Shcherbitskii in the Ukraine made his removal in 1989 a slow and complicated process requiring elaborate institutional maneuvering by Gorbachev. The protests in Kazakhstan were the costly consequence of Kunaev's removal. Nonetheless, Moscow has been willing to pay these costs in many instances; in 1988 and 1989 the first secretaries in ten of the fourteen union-republic parties were replaced. And in at least eight of these instances the secretaries were removed because of displeasure with republic policies.[70]

The cadres run a second risk: that those threatened by assertive federalism will initiate counteraction. Indeed, it was Moldavian pressures for new language legislation that ignited protests by the republic's Gagauz minority; but the agenda of the Gagauz went beyond language to include complaints of discrimination in economics and politics. The mobilization of Estonians, Latvians, Lithuanians, and Moldavians brought the countermobilization of their "exclave" Russians behind *Interdvizhenie*, the Committee for Defense of Soviet Power in Lithuania, and *Edinstvo* (Unity). In the future, similar counterprotests may be mobilized by cadres of the less developed republics of Central Asia in order to blunt the political impact of mass demonstrations in the Baltic and Caucasus against the allocation decisions made in Moscow.[71]

[68] This is not to say, as some have alleged, that the motive of the Party leaders was to instigate violence as a pretext for a military crackdown; *New York Times*, February 19, 1990. See also *Izvestiia*, January 15, 1990, January 16, 1990.

[69] *Report on the USSR* 1 (January 27, 1989), 29.

[70] *Pravda*, January 13, 1988, May 21, 1988, June 17, 1988, October 5, 1988, October 21, 1988, April 15, 1989, June 23, 1989, June 24, 1989, September 29, 1989, November 17, 1989.

[71] *Izvestiia*, March 17 1989, May 25, 1989, August 23, 1989; *Pravda*, July 29, 1989, August

A third risk to cadres is that political action mobilized by them may actually facilitate the creation of counterelites and their mobilizational activities. Much of the apparent "bandwagoning" the cadres seek to promote may actually be used as opportunities for "piggybacking" by potential counterelites; that is, aspiring independent political entrepreneurs maintain their particular agendas while using "official" political action as a cover for their own mobilizational activities and as an opportunity to publicize their agendas.[72] Rather than depriving these potential counterelites of support, the cadres' gambit may provide their opponents with the opportunity to expand it. In some republics the political action mobilized by the cadres has become self-sustaining and has slipped from their control, only then to provide counterelites with a ready-made movement. In Armenia the cadres appear to have lost control over the Karabakh protest by the late summer of 1988. According to a TASS report: "Taking advantage of the fact that the former leaders of the central committee of the Communist Party of Armenia let the initiative slip from their hands and retreated step-by-step, members of the [Karabakh] committee created ramified organizational and political structures."[73] In 1989 the initiative in Azeri protests over the status of Nagornyi Karabakh apparently passed from the cadres to counterelites within the popular front. By early 1990 the Azeri popular front was complaining that the initiative in the communal violence had even slipped from its control and passed to radicals in the National Defense Committee.[74] In Lithuania, Party First Secretary Algirdas Brazauskas was reportedly surprised when Sajudis deputies he had helped elect to the republic's Supreme Soviet rejected him as the republic's chief of state and elected their own leader, Landsbergis.[75]

Finally, the cadres' gambit may encourage unwanted primordial violence. The mass demonstrations in Erevan and Baku in 1988, 1989, and 1990 were catalysts for waves of violence across the republics. Small groups of Armenians and Azeris attacked minorities in their republics by intimidating many into emigration and by simply killing others.

10, 1989, August 11, 1989, August 23, 1989, August 25, 1989, September 5, 1989; *Sovetskaia Estoniia*, March 1, 1989, March 7, 1989.

[72] *Izvestiia*, July 13, 1989; *Komsomol'skaia Pravda*, March 3, 1989; *Sovetskaia Litva*, February 25, 1989.

[73] Quoted in Cockburn (fn. 41), 178.

[74] See *Pravda*, April 9, 1989, August 23, 1989, August 24, 1989, September 10, 1989, February 2, 1990; *Zaria Vostoka*, April 12, 1989, April 14, 1989; *New York Times*, August 28, 1989, August 29, 1989, August 30, 1989, September 1, 1989, September 17, 1989; Saulius Girnius, "Sajudis' Parliament Statement on Independence," *Report on the USSR* 1 (September 15, 1989), 17–18.

[75] *New York Times*, March 13, 1990.

POLITICAL INSTITUTIONS AND POLITICIZED ETHNICITY

The rise of assertive ethnofederalism in the Soviet Union since the mid-1950s points up how political institutions shape the mobilization of ethnic communities. On the one hand, Soviet federalism delayed the *origins* of politicized ethnicty, but on the other hand, it distributed mobilizational resources such as entrepreneurial skills and means of communications in a manner that eventually shaped its *incidence* and *agendas*.

This emphasis on institutions offers an amendment to those studies that explain the rise of politicized ethnicity by emphasizing attitudes rather than the resources needed to mobilize an ethnic community. These studies have introduced at least two alternative paradigms for the study of ethnopolitics, which are distinguished from one another by alternative views of the nature of ethnic identity, the sources of cohesion, and the objectives of politicized ethnicity and protest. The *primordialist* paradigm sees ethnic identities as one of the givens of social existence, shaped by historic memory, language, religion, and geographic compactness. The politicization of ethnicity is communal self-discovery; protest, often an expressive act affirming communal solidarity.[76] The *instrumentalist* paradigm sees ethnic identities as contingent and changing self-ascribed roles. The politicization of ethnicity and protest are goal-oriented behaviors—often focused on the pursuit of socioeconomic gain. According to the instrumentalist paradigm, an ethnic identity becomes a basis for collective action when there are comparative advantages to be gained from that specific ethnic identity over alternative ethnic, class, or other identities.[77]

The prevailing paradigm in Sovietology has been primordialist. In explaining the *origins* of Soviet ethnopolitics, it focuses our attention on attitudes such as assimilation rather than on incentives and constraints on action.[78] Certainly these attitudes are important ingredients of the

[76] Smith (fn. 7), 105. See also Walker Connor, "Nation-Building or Nation-Destroying?" *World Politics* 24 (April 1972), 319–55; Milton Esman, ed., *Ethnic Conflict in the Western World* (Ithaca, N.Y.: Cornell University Press, 1977); Clifford Geertz, "The Integrative Revolution: Primordial Sentiments and Civil Politics in the New States," in *Old Societies and New States* (New York: Free Press, 1963), 105–57; Edward Shils, "Primordial, Personal, Sacred, and Civil Ties," *British Journal of Sociology* 8 (June 1957), 130–45.

[77] Crawford Young, *The Politics of Cultural Pluralism* (Madison: University of Wisconsin Press, 1976), 43. See also Frederick Barth, ed., *Ethnic Groups and Boundaries* (Boston: Little, Brown, 1969); Glazer and Moynihan (fn. 10); Robert Melson and Howard Wolpe, *Nigeria: Modernization and the Politics of Communalism* (East Lansing: Michigan State University, 1971); Ronald Rogowski, "Causes and Varieties of Nationalism: A Rationalist Account," in Edward A. Tiryakian and Ronald Rogowski, eds., *New Nationalisms of the Developed West* (Boston: Allen and Unwin, 1985), 87–108.

[78] Arend Lijphart, "Political Theories and the Explanation of Ethnic Conflict in the Western World," in Esman (fn. 76), 46–64.

current crisis. Yet there is little indication that these alone can account for the rise of ethnopolitics—little indication that assimilation has become any less advanced, assimilative pressures any more intense, or national consciousness any higher in the decades since 1955 than in those before. The failure of Soviet policies of ethnic fusion (*sblizhenie*) to reduce cultural barriers among ethnic groups and to bring about the merger (*sliianie*) of nations cannot account for the *rise* of ethnic assertiveness over the past three decades.

In predicting the *incidence* of ethnic assertiveness, the primordial paradigm argued that the most extensive resistance to the policies of the Soviet regime would be mounted by those minorities that were culturally remote from the Russian majority.[79] In particular, it focused our attention on Islam, arguing that this provides a cultural bond among Soviet Muslims and, in the words of Kemal Karpat, "creates an invisible barrier separating them from the ruling Slavs."[80] Moreover, as Alexandre Bennigsen and S. Enders Wimbush argue, "the critical issue determining the extent and degree of long-term commitment of Soviet Muslims to the Soviet Russian state is not 'socio-economics' but identity."[81] Contrary to these expectations, however, those nationalities most remote in culture from the Slavs have been among the most quiescent. There is little evidence to suggest that it is the relative strength of their primordial sentiments that distinguishes the minorities that have engaged in sustained, effective political action from the more quiescent minorities. The primordial paradigm had not predicted the pattern of *flash points*.[82]

The primordial paradigm had also predicted that protest would mobilize behind *agendas* of cultural expression. Yet Soviet ethnofederalism has raised many issues that are poorly explained by this paradigm. For example, the recommendation by leaders of the Nagorno-Karabakh Armenians, Abkhazians, and South Osetians to transfer jurisdiction over their homelands to the Russian Republic evinces an acceptance rather than rejection of Soviet federalism and of its institutional protections for minority interests. The terms in which they have cast their legislative

[79] Teresa Rakowska-Harmstone, "The Soviet Union," in Robert G. Wirsing, ed., *Protection of Ethnic Minorities: Comparative Perspectives* (New York: Pergamon Press, 1981), 127. For a fine critical discussion of this approach, see Gail Warshofsky Lapidus, "Ethnonationalism and Political Stability: The Soviet Case," *World Politics* 36 (July 1984), 555–80.

[80] Karpat, "Moscow and the 'Muslim Question,'" *Problems of Communism* 32 (November–December 1983), 79.

[81] Bennigsen and Wimbush, *Muslims of the Soviet Empire* (London: C. Hurst, 1985), 3, 31.

[82] The paradigm might be "saved" by claiming that assertiveness of ethnic leaders in the Baltic, followed by such assertiveness in the Caucasus, is evidence of the relative levels of ethnic awareness in these lands. But in this sense the concept of relative levels of national consciousness is not a predictor of politicized ethnicity or part of a causal relationship; it is, rather, a description or definition.

agendas concerning autonomy, indigenization, language, developme and migration suggest that the public agenda for most ethnic cadres less a primordial assertion of cultural identity than an instrumental pu suit of other interests. Thus, the predictions of the prevailing paradigı concerning the *agendas* of ethnopolitics were inaccurate.

Alternatively, an analysis that draws solely upon the instrumentali paradigm would lead to significant misprediction as well. It would mis the constraint of primordialist sentiments in the population that must b co-opted into the agendas of Soviet ethnic entrepreneurs. It would dis miss the possibility that cadres might switch to primordial agendas a incentives and institutional constraints change. It would be unable t explain the willingness of some counterelites such as Landsbergis to sac rifice socioeconomic benefits for symbolic issues of self-expression.

These mispredictions point up the narrow or incomplete nature of paradigm, such as the primordialist or instrumentalist, that focuses o attitudes. These paradigms fail to take two facts into consideration. First the attitudes that sustain either primordial or instrumental agendas exis side by side in many Soviet ethnic communities and often within th same individual.[83] They coexist among Soviet cadres as well as in th general population. Second, the politicization of either primordial senti ments or instrumental interests and the mobilization of ethnic commu nities in sustained, large-scale action has required the conjunction o these attitudes with resources that can mobilize an ethnic community The attitudes cited by one or the other paradigm are necessary, but no sufficient, for the explanation of ethnopolitics.

Political institutions like Soviet federalism play a critical role in thi conjunction and so in shaping ethnic communities, politicizing ethnicity and mobilizing protest. They empower entrepreneurs and constrai their choice of either primordial or instrumental strategies.[84] The politi

[83] Suny (fn. 63), 220. This may also be true in many Western societies; see Anthon Mughan, "Modernization and Ethnic Conflict in Belgium," *Political Studies* 27 (March 1979) 21–37; Rogowski, "Conclusion," in Tiryakian and Rogowski (fn. 77), 374–76.

[84] Karl W. Deutsch has noted that political systems also differ in the degree to which thei institutions influence ethnic attitudes (particularly assimilation); see Deutsch, *Political Com munity at the International Level: Problems of Definition and Measurement* (Garden City, N.Y. Doubleday, 1954), 33, 39–40. The "new institutionalism" argues a more general case for th "endogeneity" of preferences; for example, James P. March and Johan P. Olsen argue tha traditionally "political theory has treated political institutions as determining, ordering, o modifying individual motives"; see March and Olsen, *Rediscovering Institutions: The Organi zational Basis of Politics* (New York: Free Press, 1989), 4–7, 154–56. For additional motive that incline ethnic elites toward intercommunity conflict regulation, see Eric A. Nordlinger *Conflict Regulation in Divided Societies* (Cambridge: Harvard University Center for Interna tional Affairs, 1972).

cization of ethnicity has been the work of political entrepreneurs created by Soviet federalism.[85] In the three decades before 1990 the most significant pressure on Moscow for ethnic interests was orchestrated by the Party and state leadership of ethnic homelands. By assigning a monopoly over mobilizational resources, Soviet federalism delayed the rise of ethnofederalism but at the same time made it likely and possible that ethnic cadres of more developed ethnic communities would later mobilize their constituents. By constraining their choice of strategies, Soviet federalism made it likely that cadres would mobilize their constituents behind instrumental rather than primordial agendas. Since January 1990 the magnitude, patterns, and agendas of ethnic assertiveness have evolved most in those areas in which institutional changes have transformed the incentive structures of cadres or empowered new elites (notably in the Baltic republics).[86]

Where political institutions neither establish monopolistic ethnic entrepreneurs nor constrain their choice of strategies in this way, political institutions distribute mobilizational resources and shape ethnic strategies differently.[87] This is illustrated poignantly by the changes in Soviet political institutions. In some republics the cadres' control over their communities has been weakened by the loss of their monopoly over the mobilizational resources of the ethnic community. The policies of *demokratizatsiia* and *glasnost* have undermined the ability of cadres to contain the protest they have mobilized as well as to suppress autonomous ethnopolitics. Where threats to their control have grown, their strategies have often changed. Cadres in some republics now engage in competitive efforts to mobilize different segments of elite and popular constituencies behind competing agendas reflecting different balances of primordial and instrumental concerns.[88] By early 1990 competition had led to an outright split in the Communist parties of the republics of Latvia and

[85] Illustrations of this approach include Edward Allworth, "Restating the Soviet Nationality Question," in Allworth, ed., *Soviet Nationality Problems* (New York: Columbia University Press, 1971); Helene Carrere d'Encausse, *Decline of an Empire* (New York: Newsweek Books, 1978); Rakowska-Harmstone (fn. 79).

[86] Gorbachev's initial reponse to new elites such as Landsbergis and his Sajudis government has been consistent with the Soviet ethnic strategy that has emphasized institutional constraints to shape ethnic agendas. Using deterrent and compellent constraints, he has sought to induce the Lithuanian government to abandon its primordial agenda and pursue the instrumental objectives of *perestroika* within the context of Soviet federalism.

[87] Sidney Tarrow described how in democratic systems as well the strategies of regional leaders acting as brokers between center and periphery are shaped by bureaucratic institutions; see Tarrow, *Between Center and Periphery: Grassroots Politics in Italy and France* (New Haven: Yale University Press, 1977), 7–8, 43–44.

[88] *Sovetskaia Latviia*, March 19, 1989.

Lithuania. Alternative leaderships have articulated competing ethnic agendas, and Party assets including buildings have been divided.[89]

Lithuania before and after January 1990 provides an excellent diachronic study of how changes in the Soviet electoral system can empower new political entrepreneurs in the homeland and shift the emphasis in ethnic agendas from instrumental to primordial issues. The pragmatic leadership of Brazauskas was replaced by that of Landsbergis, which immediately emphasized symbolic issues of sovereignty even if they did not impart real autonomy. Pressed by Moscow, the new leadership quickly agreed that all instrumental issues were negotiable but not the primordial ones—that is, the symbolic declaration of independence.[90] Because Soviet institutions have not changed evenly in all republics, the contrast between Lithuania and certain other union republics after January 1990 offers a cross-sectional comparison of the consequences of changed institutions. In Belorussia, for example, the elections of early 1990 did not offer an open contest of elites and failed to empower alternative political entrepreneurs. The Belorussian political agenda remained much more instrumental and supportive of Moscow, so that in Moscow's confrontation with the Sajudis government, the Belorussian leadership announced it would demand renegotiation of its borders with Lithuania should the latter secede from the Soviet Union.[91]

The Soviet experience with ethnofederalism illustrates the importance of bringing institutions back into the analysis of ethnopolitics—for Sovietology, for comparisons of Leninist with non-Leninist polities, and possibly for comparisons of Soviet politics before and after 1990. Within the institutions of Soviet federalism may be keys to the future of the Soviet system. In particular, the direction the Soviet Union takes will depend on the ability of ethnic cadres to adapt to the role of entrepreneurs in a competitive arena as well as the ability of Soviet institutions to constrain the cadres' choice of strategies to those that do not threaten the unity of the Soviet polity.

[89] *Pravda*, February 8, 1990, February 9, 1990: *New York Times*, March 24, 1990. In Lithuania, sensing the consequences of these institutional changes for their power, members of the republic Politburo, such as the prime minister of the Sajudis government, have also begun to resign from the Communist Party.

[90] *New York Times*, March 30, 1990.

[91] *New York Times*, March 5, 1990, March 6, 1990, March 31, 1990.

EXPLAINING THIRD WORLD ALIGNMENT

By STEVEN R. DAVID*

MOST of the earth's people and states are part of the Third World. Yet Third World foreign policy, especially why Third World states align as they do, is not well understood. Balance of power, the most widely used theory in international relations, is particularly inadequate as an explanation of Third World alignments because it does not take into account the distinctive characteristics of the Third World. The need for a theory that considers those characteristics is critical because understanding the Third World is and will remain central to understanding the course of international politics.

The theory of "omnibalancing" meets this need. It draws upon some of the key assumptions of balance of power while also correcting those elements of the theory that make it inapplicable to the Third World. Omnibalancing agrees with the central assumption of balance of power—that threats will be resisted. But it departs from balance of power in explaining Third World alignment decisions as a result of the Third World leadership's need to counter *all* threats. Thus, whereas balance of power focuses on the state's need to counter threats from other states, omnibalancing considers internal and external threats to the leadership, and, as a result, it fundamentally alters our understanding of why Third World leaders align as they do and also provides insights that explain a wide range of Third World behavior.

The argument is presented in four parts. First, the theories of balance of power and omnibalancing are reviewed, with special emphasis on the points of agreement and disagreement. Second, the characteristics that distinguish the Third World from the rest of the world and why these characteristics support the applicability of omnibalancing over balance of power are considered. Third, two specific cases are examined. These illustrate how omnibalancing explains and predicts Third World alignments in ways that balance of power (as well as other theories) do not. Finally, there is a discussion of the centrality of an understanding of the Third World, including Third World alignment, to the study of international politics.

* Of the many who offered advice, I would like to thank especially Aaron Friedberg and Stephen Van Evera for their detailed comments and suggestions.

Balance of Power

Alignment occurs when a state brings its policies into close cooperation with another state in order to achieve mutual security goals.[1] Balance of power theory argues that states align to protect themselves against the power of or threats from other states.[2] The key concern of states is to align in such a way as to prevent (by balancing against) any other state or group of states from achieving preponderance. States behave this way because they recognize that the emergence of a hegemonic power will ultimately threaten their own survival. Balance of power theory emphasizes that the determinants of alignment come overwhelmingly from the structure of the international system, particularly the actual and potential *external* threats that states face.[3] Accordingly, the internal characteristics of the states are usually not considered relevant in influencing alignments.[4]

This view is reinforced by the recurring formation of balances made up of states that have little more than a threat in common. The alliance of democratic France with tsarist Russia against Germany prior to World War I illustrates this.[5] The apparent irrelevance of internal factors is reinforced by the fact that states with similar political systems that might otherwise be expected to maintain an alignment do not in the absence of a common external threat.[6] An example of this phenomenon

[1] For a comprehensive discussion of what connotes alignment, see George T. Duncan and Randolph M. Siverson, "Flexibility of Alliance Partner Choice in a Multipolar System," *International Studies Quarterly* 26 (December 1982), 511–38, at 518.

[2] In my use of the term balance of power, I include the efforts of statesmen to counter both power and threats. For more on this point, see Stephen M. Walt, *The Origins of Alliances* (Ithaca, N.Y.: Cornell University Press, 1987), 5, 21–26; and Inis L. Claude, Jr., *Power and International Relations* (New York: Random House, 1962), 64–65.

[3] Kenneth Waltz and, to a lesser extent, Hans Morgenthau base their balance of power theory on the capabilities of states. Generally speaking, the greater the capability of a state, the greater the threat it poses to other states; see Waltz, *Theory of International Politics* (Reading, Mass.: Addison-Wesley, 1979), esp. chap. 6; and Hans Morgenthau and Kenneth Thompson, *Politics among Nations*, 6th ed. (New York: Knopf, 1985), esp. chap. 11. George Liska also emphasizes the role of external threat, calling it the "primary source of alliances"; see *Nations in Alliance: The Limits of Interdependence* (Baltimore, Md.: Johns Hopkins University Press, 1962), 13.

[4] Ole R. Holsti, P. Terrence Hopmann, and John D. Sullivan, *Unity and Disintegration in International Alliances: Comparative Studies* (New York: John Wiley, 1973), 5. Based on an exhaustive survey of the literature, the authors concluded that most balance of power writings argue this.

[5] Kenneth Waltz, "Theory of International Relations," in Fred I. Greenstein and Nelson W. Polsby, eds., *Handbook of Political Science: International Politics* (Reading, Mass: Addison-Wesley, 1975), 8:43.

[6] The notion is widespread that alliance cohesion weakens when there is no external threat. See, for example, Arnold Wolfers, *Discord and Collaboration* (Baltimore, Md.: Johns Hopkins University Press, 1962), 29; and Kenneth E. Boulding, *Conflict and Defense: A General Theory* (New York: Harper and Row, 1962), 162.

ailure of the Soviet Union and the People's Republic of China to aligned despite their common allegiance to communism.

ough they base their work on the experience of states outside the World, many balance of power theorists argue that their theory or the states of the Third World as well.[7] One can expect the of power theory to prevail, it is argued, whenever states seeking ive and expand their power interact in an anarchic order.[8] Since nothing in the nature of Third World states or in the internaystem they inhabit that would contradict these requirements, balpower politics should apply equally to them.

Omnibalancing

gh the balance of power theory contains important insights, it ot adequately explain Third World alignment and realignment. , the most powerful determinant of Third World alignment beis the rational calculation of Third World leaders as to which power is most likely to do what is necessary to keep them in This is so because of the unstable, dangerous, and often fatal of the international and *domestic* political environment that chars the Third World. This deceptively simple observation changes lerstanding of why Third World leaders align and realign, and it s those decisions better than does balance of power. The uniquely ous context in which Third World politics takes place also exwhy theories derived from the eighteenth- and nineteenth-century an experience (such as balance of power) are inadequate in exg the alignment decisions of Third World leaders.

essence of balance of power theory is correct: leaders of Third states as well as leaders of other states align in ways that help etter resist the threats they face. But conditions in the Third require, in addition, a theory of alignment that applies primarily Third World. To that end, three separate and distinct repairs of of power theory are offered. First, rather than just balance threats or power, leaders of states will appease—that is, align secondary adversaries so that they can focus their resources on dversaries. Other states may be less threatening than the secondat, but it may be that they cannot be harnessed for the alignment

s a major argument made by Stephen Walt, who asserts that his theory is supported eriences of states in the Middle East. He does, however, modify balance of power focus on threats (instead of power) from other states. See Walt (fn. 2), 13–14.
: (fn. 3), 121; and Morgenthau and Thompson (fn. 3), 228.

either because they do not care about the dispute or because they ar
unable to resolve it. So the threatened leadership has no choice: it mus
align with one threat to address the other. The Soviet Union's alignmen
with the United States and Great Britain against Nazi Germany durin
World War II is an example of this type of behavior.

As with the leaders of great powers, Third World leaders, too, seek t
appease secondary threats in order to counter those that are more press
ing. But in the Third World, this often means appeasing other state
(which often pose less pressing threats) in order to counter the mor
immediate and dangerous domestic threats. They seek to split the align
ment against them and focus their energies on their most dangerous (do
mestic) opponents. To do this they appease the international allies o
their domestic opponents. This superficially appears to be "bandwagon
ing," that is, a case of a state accommodating a threatening adversary i
the hope of appeasing it or sharing in the spoils.[9] But it is better classifie
as balancing; the accommodation, made to conserve strength for the bat-
tle against the prime threat, is part of a general policy of resistance.

Finally, since the dominant goal of Third World leaders is to stay i
power, they will sometimes protect themselves at the expense of the in-
terests of the state. This form of balancing behavior—not covered b
balance of power theory—must be addressed if the alignment decision
are to be understood.

Omnibalancing, the theory that emerges from these observations, ap-
plies largely to the Third World. It incorporates the need of leaders t
appease secondary adversaries, as well as to balance against both interna
and external threats in order to survive in power. This theory rests o
the assumptions that leaders are weak and illegitimate and that th
stakes for domestic politics are very high—conditions that are much
more common in the Third World than elsewhere. It assumes that th
most powerful determinant of alignment is the drive of Third World
leaders to ensure their political and physical survival.

Although omnibalancing is framed in opposition to balance of power
it also shares much of the latter's realist underpinnings. In accordanc
with realism, omnibalancing argues that international politics focuses on
power, interests, and rationality.[10] Omnibalancing accepts the realist
view that the world is characterized by international anarchy, that inter-
ests are bound to conflict, and that the use of force is always an option

[9] The first mention I have been able to find of "bandwagoning" in the sense used here i
in Arnold Wolfers, *Discord and Collaboration: Essays in International Politics* (Baltimore, Md.
Johns Hopkins University Press, 1962), 15. See also Walt (fn. 2), 19–21.

[10] See Morgenthau and Thompson (fn. 3), chaps. 1, 3, for a concise description of realism.

for heads of state. Omnibalancing furthermore accepts the realist argument that there are clear hierarchies of issues, with survival being the most important. In addition, omnibalancing rests on the fundamental realist assumption that "politics, like society in general, is governed by objective laws that have their roots in human nature," namely, in this instance, the desire of Third World leaders to survive.[11]

Omnibalancing also acknowledges the central lesson of realism, that international politics can best be understood by asking an observer to imagine how she or he would rationally behave under given conditions.[12] A decision is elucidated, therefore, by showing how it was reasonable in light of the leader's objectives.[13] As Hans Morgenthau explains:

> In other words, we put ourselves in the position of a statesman who must meet a certain problem of foreign policy under certain circumstances, and we ask ourselves what the rational alternatives are from which a statesman may choose who must meet this problem under these circumstances (presuming always that he acts in a rational manner), and which of these rational alternatives this particular statesman, acting under these circumstances, is likely to choose.[14]

Despite the overall realist approach, omnibalancing departs in fundamental ways from the understanding of realism put forth by scholars such as Morgenthau and Waltz.[15] The essence of the difference is that omnibalancing asserts that realism must be broadened to examine internal threats in addition to focusing on external threats and capabilities (that is, structural arguments), and that the leader of the state rather than the state itself should be used as the level of analysis. As such, omnibalancing challenges the core assumptions of Morgenthau and Waltz, that states are unitary actors pursuing the national interest. If omnibalancing is correct, then alignment decisions of Third World leaders cannot be understood without reference to the role of internal threats to their leadership.

This approach changes the way we go about explaining alignment

[11] Ibid., 4.

[12] This lesson comes from the two central works of realism, Morgenthau's *Politics among Nations* (fn. 3) and Thucydides' *History of the Peloponnesian War* (New York: Penguin, 1985). See also Robert O. Keohane, "Theory of World Politics: Structural Realism and Beyond," in Keohane, ed., *Neorealism and Its Critics* (New York: Columbia University Press, 1986), 163, for elaboration of this point.

[13] This is in conformity with Graham Allison's Model I; see Allison, *The Essence of Decisions: Explaining the Cuban Missile Crisis* (Boston: Little Brown, 1971), 5.

[14] Morgenthau and Thompson (fn. 3), 5.

[15] See, for example, Morgenthau (fn. 3), esp. chaps. 1, 3, as well as Waltz (fn. 3). Although differences exist between the two interpretations, they are treated together in that they both accept the view of realism that envisions a world of international anarchy in which the principal actors are states that seek to expand and survive.

decisions. Balance of power assumes that the decision maker asks, "How does this policy affect the power of the state?"[16] Omnibalancing assumes instead that the decision maker asks, "How does this policy affect the probability of *my* remaining in power?" And whereas balance of power assumes the state's leaders ask, "Which outside power is most likely to protect my state from the threats posed by other states?" omnibalancing assumes the decision maker asks, "Which outside power is most likely to protect *me* from the internal and external threats (as well as combinations of both) that I face." Just as the questions differ, so do the responses.

By taking internal threats into account, omnibalancing is less parsimonious than is balance of power in explaining Third World alignment and realignment. The loss in simplicity is more than compensated, however, by the increased explanatory and predictive power of the theory. Demonstrating this requires showing that one can generalize about the Third World and that these generalizations support the view that a broader conception of balance of power is needed in order to explain why Third World leaders make the alignment decisions they do.

Generalizing about the Third World

Despite the vast differences among the states considered as belonging to the Third World, there are also fundamental similarities that justify considering them together.[17] These similarities have created a situation in which internal threats (with or without external backing) are far more likely to challenge a Third World leader's hold on power than are threats from other states. Only a handful of Third World leaders (for example, Uganda's Idi Amin, Cambodia's Pol Pot, Panama's Manuel Noriega) have fallen to outside invasions. Hundreds of Third World leaders, however, have been overthrown by their internal enemies. In fact, most Third World states have experienced either a forceful overthrow of their government (from within) or a nearly successful attempt. Since 1945 wars within states have outnumbered wars between states, a gap that is especially pronounced in the Third World.[18] Coups d'état alone have

[16] Morgenthau and Thompson (fn. 3), 14.

[17] This study considers the Third World as including all countries *except* the United States, the Soviet Union, Canada, Japan, Australia, New Zealand, South Africa, the European states, and the People's Republic of China. For an especially good description of the differences between Third World states, see Benjamin Cohen, *The Question of Imperialism: The Political Economy of Dominance and Dependence* (New York: Basic Books, 1973), 145–52.

[18] J. David Singer and Melvin Small, *Resort to Arms: International and Civil Wars, 1816–1980* (Beverly Hills, Calif.: Sage, 1982), 92–95, 98–99, 229–32. For additional treatment of this

accounted for nearly two hundred regime changes in the Third World, making it the most common form of coercive regime change and perhaps the most common form of regime change in general in the Third World.[19] Third World leaders understandably have to pay special attention to retaining their hold on power, especially in the face of domestic challenges.

The reasons for the prevalence of internal threats in the Third World stem from the characteristics of most Third World states.[20] Unlike countries outside the Third World that evolved over centuries, the great majority of Third World states had been colonies out of which foreign powers created states where none had existed. Although the degree to which the newly formed boundaries coincide with the boundaries of indigenous societies varies in the Third World (for example, it is high in Southeast Asia and low in Africa), in all cases a formal division replaced what had been a flexible demarcation. Because of the arbitrariness of their borders, therefore, many Third World states began and remain more as an artificial construct than a coherent unit.[21]

The artificiality of the Third World state has created a situation in which subnational groups owe allegiance to and act on behalf of interests other than the national interest. Instead of identifying with the state, individuals identify with ethnic, religious, or regional groupings. This type of narrowly defined interest perpetuates itself by preventing the formation of a national consciousness. With anticolonial nationalism dissipating in the wake of decolonization, identification with the group over the state is getting stronger with time.[22] Rather than transcending the differences among these different groups, the state is often simply the representative of a group that holds power in the capital. In such a situ-

---

point, see John Lewis Gaddis, *The Long Peace: Inquiries into the History of the Cold War* (New York: Oxford University Press, 1987), 224.

[19] From 1945 to mid-1985 there were at least 183 successful coups and 174 unsuccessful coup attempts in the Third World. See Steven R. David, *Third World Coups d'Etat and International Security* (Baltimore, Md.: Johns Hopkins University Press, 1987), 1–2. These numbers are just conservative estimates; the actual number of coups and coup attempts may be much higher. For a comprehensive examination of estimates of numbers of coups, see Milton Leitenberg, "Appendix 2: World-Wide Military Coups since 1945: A Short Note on Data Collection," in Asbjorn Eide and Marek Thee, eds., *Problems of Contemporary Militarism* (London: Croom Helm, 1980), 378–85.

[20] For an excellent discussion of why internal threats are so common in the Third World, see Mohammed Ayoob, "Security in the Third World: The Worm about to Turn?" *International Affairs* 60 (Winter 1983–84) 41–52.

[21] Robert H. Jackson and Carl G. Rosberg, "Why Africa's Weak States Persist: The Empirical and the Juridical in Statehood," *World Politics* 35 (October 1982), 1–24; Mutiah Alagappa, *The National Security of Developing States: Lessons from Thailand* (Dover, Mass.: Auburn House, 1987), 4–6; Christopher Clapham, *Third World Politics* (Madison: University of Wisconsin Press, 1985), 8.

[22] Caroline Thomas, *In Search of Security* (Boulder, Colo.: Rienner, 1987), 2.

ation, the leadership becomes just one other contender in a struggle for power that knows few constraints.[23]

Legitimacy is likely to be weaker for Third World leaders than for leaders elsewhere. Many regimes in the Third World are narrowly based, came to power through force, and use suppression to remain in power. Because they lack legitimacy, they face continual threats to their rule.[24] Threats are also likely to emerge because Third World leaders control much more wealth and power than do the other groups in the society. Gaining control of the state is therefore the only means by which the ambitious can meet their needs. Hence, the state is especially vulnerable because many different groups and individuals seek to control it. At the same time, those in power will mightily resist attempts to replace them because they do not want to relinquish their only opportunity to acquire and keep wealth and influence.[25] Equally important, since the consequence of loss of power in the Third World is often loss of life, Third World leaders are understandably more aggressive than other leaders in their efforts to maintain their positions.

Most Third World states are governed by authoritarian rule, which exacerbates and focuses the conflicts that do arise. Despite the rise of nationalism, meaningful political participation by the population at large in most Third World countries remains the exception. Decisions, especially in foreign policy, are usually made by a single individual or at most by a narrow elite. Although this elite does not have total flexibility in making policy, it typically needs to be far less responsive to national or institutional concerns than leaders elsewhere need to be. Public opinion in Third World states often carries little influence, and bureaucracies play only a minor role in foreign policy.[26] Moreover, due to the effects of colonialism and cultural factors, rulers see themselves and are seen as an alien imposition on the society.[27]

An interrelationship between internal and external threats also characterizes the Third World. Third World leaders and insurgents commonly seek outside support to advance their interests. Their requests for help either to threaten or to protect a state's leadership are often granted

[23] For an excellent discussion of the role of subnational groups in Third World states (in this case, West Africa), see I. William Zartman, *International Relations in the New Africa* (Englewood Cliffs, N.J.: Prentice-Hall, 1966), 47–48.

[24] Alagappa (fn. 21), 9. The problems of weak legitimacy are as common to the "old" Third World states of Latin America as they are to the newer states of Africa and Southeast Asia.

[25] Clapham (fn. 21), 39–43.

[26] Joel S. Migdal, "Internal Structure and External Behaviour: Explaining Foreign Policies of Third World States," *International Relations* 4 (May 1974), 519–20.

[27] Clapham (fn. 21), 19.

by outside states.[28] Internal threats, so prevalent in the Third World, are an ideal vehicle for advancing the interests of outside states. Since the leadership of Third World states is often determined by the outcome of internal threats, other countries with an interest in the Third World state may seek to determine that outcome. Moreover, the military weakness of most Third World states enhances their willingness to back or help suppress internal threats against their neighbors. Most Third World states simply lack the weapons and the logistical capability for direct, protracted conflict beyond their borders. Although wars between states do occur, and although troops may at times be sent far afield (as in the case of Cuba), these actions are relatively rare. At the same time, the experience of Third World states that have employed external attacks (for example, Egypt and Syria versus Israel in 1967 and 1973; the Iran-Iraq War of 1980–88) confirms the prohibitive price and risks incurred by states that mount wars against formidable adversaries. Providing aid to or against insurgent groups, on the one hand, or backing or suppressing coups, on the other hand, is a relatively inexpensive and effective means of asserting one's power.[29]

Finally, one can categorize Third World states according to their self-identification. A country that considers itself to be a Third World country is likely to have a set of attitudes and goals that will be defined by its "Third Worldness." This is not a tautology. Third World states overwhelmingly supported the huge price increases in OPEC oil despite the devastating economic impact on them. Third World solidarity has also been displayed in the Non-aligned Movement and the Law of the Sea Conference. One can speak of a Third World, therefore, because member states do and because they act in some ways at least because of that self-identification.[30]

[28] On foreign involvement in internal conflicts, see James Rosenau, ed., *International Aspects of Civil Strife* (Princeton: Princeton University Press, 1964); Asti Suhrke and Lela N. Garner, eds., *Ethnic Conflict in International Relations* (New York: Praeger, 1977); and Herbert K. Tillema, "Foreign Overt Intervention in the Nuclear Age," *Journal of Peace Research* 26 (1989), 179–85.

[29] Even threats such as coups d'état, which are thought of as strictly internal, frequently have a foreign dimension. From 1945 to mid-1985 foreign involvement played a significant role in the support of at least twenty-four successful and unsuccessful coup attempts and in efforts to suppress fourteen coups in the Third World. Foreign involvement was thus involved in slightly over 10% of the total of all coups and coup attempts. See David (fn. 19), 2.

[30] According to Stephen Krasner, Third World states behave alike in global negotiations, not out of a common desire for economic development, but rather out of a common need for regime security. This view supports the contentions that one can generalize about the Third World and that concerns over survival assume central importance in Third World decision making. See Krasner, "Third World Vulnerabilities and Global Negotiations," *Review of International Studies* 9 (October 1983), 235–51.

These generalizations about the Third World are not intended to suggest that all Third World countries share these characteristics equally. Different states manifest different strengths and weaknesses. Nor do these generalizations apply only to the Third World. As clearly demonstrated by the toppling of Eastern European governments in 1989, states outside of the Third World also suffer from such problems as lack of consensus and weak legitimacy. What justifies considering the Third World as an analytic category is that whatever combination of factors may exist in a particular Third World state, the cumulative impact of those factors makes most Third World leaders more vulnerable to overthrow—particularly from internal threats—than leaders elsewhere. Thus, while the leaders of all states must be concerned about threats, in the Third World such concerns assume an urgency and priority that is rarely matched elsewhere.

Why the Nature of the Third World Supports the Theory of Omnibalancing

The characteristics of Third World states demonstrate why structural theories of alignment such as balance of power are inapplicable to the Third World. At the heart of balance of power theory is the distinction between the anarchy of the international system and the order of domestic politics. Adherents emphasize that international politics has no central government with the power and authority to resolve conflicts. The absence of a strong consensus in international politics, moreover, exacerbates the ensuing struggle for power. In such an environment, balancing plays a central role in ensuring the survival of states against the aggressive designs of other states. Within states, however, balance of power advocates argue that conflicts take place in an "integrated society, kept together by a strong consensus and the normally unchallengeable power of a central government."[31] Balance of power theory acknowledges that internal balancing takes place as groups combine (or balance) to advance their interests. Because the central government ensures that the survival of losers in domestic struggles is not threatened, however, supporters of balance of power assert that balancing is of much less importance in domestic politics than it is in international politics.[32]

The problem with this distinction is that central governments in the Third World often lack the power to resolve disputes within their bor-

[31] Morgenthau and Thompson (fn. 3), 190.

[32] See ibid., 189–92, for a concise description on why internal balancing is supposedly fundamentally different from balancing between states.

ders. Thus, there is often no "strong consensus" or "integrated society" to inhibit conflict, and the Third World reproduces rather than provides a haven from the anarchy of international politics. Instead of sharply differentiating what occurs on the international scene from what occurs within states, it is more accurate to consider Third World domestic politics as a microcosm of international politics. As a result, balancing to ensure survival is as critical for groups within states as it is between states.[33]

Balance of power theory focuses on the state as the principal level of analysis; that is, for most of its adherents, the state, and not the individual ruler, is the main actor in international politics.[34] They view the state as a fully formed unit with clearly defined borders capable of operating autonomously because they perceive domestic politics as "hierarchically ordered"; that is, the central government can command the obedience of different groups under its authority and thereby control what goes on within its borders. According to the theory, this justifies a focus on the behavior of the state.[35]

These assumptions, however, are of little relevance to Third World states, with their arbitrary borders encompassing many groups that often successfully resist domination from the top. It is the leadership of the state and not the state itself that is the proper unit of analysis for understanding Third World foreign policy, particularly alignment. When we refer to Egypt or Ethiopia becoming pro-American or pro-Soviet, we are really referring to the orientation of a narrow authoritarian elite. The distinction is fundamental, because the interests of this elite need not be synonymous with the interests of the state. Understanding why leaders make the alignment decisions they do requires an understanding of what is in the best interests of the leaders and not of what is in the best interests of the state. This is especially true when the survival of the leadership is at stake. Very few Third world leaders need to worry about losing their country—the mortality rate of even Third World states is very low. It is reasonable to assume, however, that Third World leaders do worry about the much more common, but (from their perspective) equally consequential, prospect of being removed from power. It follows that when a leadership is confronted with a choice between aligning so as to benefit the state but endangering its hold on power or aligning in such a way that harms the state but preserves its power, it will choose the latter.

[33] For an intriguing argument that order is more common between Third World (in this case, African) states than within them, see Jackson and Rosberg (fn. 21).

[34] For an argument of why states should be considered the main actors in international politics, see Waltz (fn. 3), 93–95; and see also Keohane (fn. 12), 160.

[35] Waltz (fn. 3), 81.

Balance of power theory assumes that states will seek to expand their power as much as they can to provide a "margin of safety" to ensure their survival.[36] In the Third World, however, the prevalence of threats against leaders places a premium, not on expanding power per se, but on surviving. In this respect, politics in the Third World supports some of the observations of neorealism, in particular, that elites have their own interests, the most important of which is to stay in power. Neorealism also recognizes that efforts to maximize power produce resistance that will ultimately jeopardize one's security. Consequently, leaders prefer to align with states that ensure their hold on power rather than with states that may increase their power, but at the risk of endangering their survival. Conditions in the Third World, however, call into question neorealism's emphasis on countering threats from other states as the prime determinant of alignment.[37]

Balance of power theory is flawed in its application to the Third World because it ignores internal threats; that is, it overlooks the most likely source of challenge to the leadership of Third World states. Balance of power theory therefore also cannot accurately assess the strength of external threats because such threats usually depend for their effectiveness on internal conditions that are susceptible to outside manipulation. In 1971, for example, the Pakistani leadership believed it had to counter a secessionist threat from the East Bengalis in order to remain in power. This meant confronting India, which supported the secessionist movement. India alone posed a considerable but manageable external threat to Pakistan. What made the Indian threat unmanageable and led to the decisive Pakistani defeat in the 1971 Indo-Pakistani War was Indian power *combined* with the massive internal threat mounted by the East Bengalis.[38] Even where there is a pressing external threat, therefore, a Third World leader's estimation of the need for outside support to counterbalance the threat often depends on internal factors that can affect the magnitude of the external threat.

In sum, in seeking to explain Third World alignment, the overall nature of the Third World supports omnibalancing while making the ap-

[36] Morgenthau and Thompson (fn. 3), 227–28.

[37] For a thorough examination of neorealism, see Keohane, ed., *Neorealism and Its Critics* (fn. 12); see also Robert O. Keohane, "Alliances, Threats, and the Uses of Neorealism," *International Security* 13 (Summer 1988), 169–76, at 173.

[38] Ayoob (fn. 20), 43. On the background of the 1971 Indo-Pakistani War (including Indian support of the East Bengalis), see Robert Jackson, *South Asian Crisis: India, Pakistan, and Bangladesh: A Political and Historical Analysis of the 1971 War* (New York: Praeger, 1975); and Pran Chopra, *India's Second Liberation* (Cambridge: MIT Press, 1974). On the need of the Pakistani leadership to prevent the secession of East Pakistan in order to survive in power, see Jackson, 26–27.

plicability of balance of power highly suspect. For those cases in which the principal threats to the leadership are from other states, balance of power theory indeed applies. But it is only prudent to consider the impact of domestic challenges before assuming that external threats alone have determined an alignment, because in what appears to be the vast majority of Third World cases, internal threats play a major role. Omnibalancing therefore needs to be invoked. Omnibalancing also respects the complexity and multifaceted nature of the threats confronting Third World leaders; it recognizes that they typically face many threats from many sources and that lesser challenges must be appeased in order to resist the principal ones. Omnibalancing is thus not misled into concluding that bandwagoning to threatening states necessarily means that the leadership chose not to balance against the most pressing threats it faced.

Cases of Omnibalancing: Mengistu and Sadat

The superiority of omnibalancing for explaining Third World alignment can be seen in the decisions of Ethiopia's Mengistu Haile Mariam to align with the Soviet Union and of Egypt's Anwar Sadat to align with the United States. These examples, of course, do not confirm the overall applicability of omnibalancing to the Third World (only an in-depth analysis of all Third World alignment decisions could do that). Nevertheless, they are particularly valuable in showing how conditions that are prevalent in the Third World prevent balance of power theory from explaining or predicting the alignments of leaders in key Third World countries. Moreover, these cases illustrate how alternative explanations of alignment, such as bandwagoning and ideology, also fail to match the ability of omnibalancing to explain the alignment decisions of these leaders.

By the standards of balance of power, the military officers (subsequently labeled the Derg) who overthrew Haile Selassie in 1974 should have made every effort to maintain Ethiopia's alignment with the United States.[39] For the twenty years prior to the Derg's takeover, the United States had been the principal supporter of Ethiopia, providing it with the largest amount of military and economic assistance sent to a sub-Saharan

[39] For background on why Mengistu aligned with the Soviet Union, see Marina Ottaway, *Soviet and American Influence in the Horn of Africa* (New York: Praeger, 1983); David A. Korn, *Ethiopia, the United States, and the Soviet Union* (Carbondale: Southern Illinois University Press, 1986); Colin Legum and Bill Lee, *The Horn of Africa in Continuing Crisis* (New York: Africana, 1979); and Paul B. Henze, *Russians and the Horn: Opportunism and the Long View*, European-American Institute for Security Research, The EAI Papers, No. 5 (Marina del Ray, Calif.: European American Institute, 1983).

African country. Moreover, despite the radical rhetoric of the Derg, . pecially that of its emerging leader, Mengistu Haile Mariam, the Unit States emphasized its intention to support the new government. To tl end, American assistance to Mengistu's regime dramatically *increa* from levels provided Haile Selassie. Fully one-third of all military equi ment delivered to Ethiopia by the United States arrived in the two yea after the Derg took power. American economic aid to Ethiopia bo bilaterally and through multilateral institutions also showed marked i creases.[40]

Any doubts about Mengistu's maintaining his alignment with tl United States ought to have (by balance of power logic) been erased b the Soviet backing of Ethiopia's chief adversary, Somalia. Because of Sc viet support, Somalia, which maintained irredentist aims against Ethic pia, had been transformed into one of the most powerful countries i sub-Saharan Africa. Beginning in 1962 and dramatically escalating afte the Derg took power in Ethiopia, Moscow provided huge amounts o military equipment and hundreds of advisers to the Somali armed forces America's increased assistance to the Mengistu government proved es sential in maintaining a rough parity between the Ethiopian and Somal forces and deterring additional Soviet involvement. It would appear tha Mengistu would do all that he could to keep American support to enabl him to continue to resist (or balance against) his Soviet-backed adversary

And yet the Derg under Mengistu chose not to intensify its alignmen with the United States but instead attempted to realign with the Sovie Union—the backer of Ethiopia's primary external threat. From the tim that Haile Selassie was overthrown, Mengistu and his colleagues on th Derg pressed the Soviet Union to replace the United States as Ethiopia's principal supporter. While still dependent on the United States for security assistance, Mengistu transformed Ethiopia along Marxist-Leninis lines and followed the Soviet lead in foreign policy. After three years the USSR finally acceded to Mengistu's requests: in 1977 the Soviet Union became Ethiopia's primary military supplier, and in 1978 a formal treaty of friendship and cooperation with the Soviet Union replaced the mutual defense treaty Ethiopia had concluded with the United States.

Why did this take place, in violation of balance of power theory? Mengistu's ideological leanings may have made him more comfortable with

[40] Korn (fn. 39), 8, 13; Bereket Habte Selassie, *Conflict and Intervention in the Horn of Africa* (New York: Monthly Review Press, 1980), 138; U.S. Arms Control and Disarmament Agency, *World Military Expenditures and Arms Transfers, 1968–1977* (Washington, D.C.: G.P.O., 1979), 157; Agency for International Development, *U.S. Overseas Loans and Grants and Assistance from International Organizations* (Washington, D.C.: G.P.O., 1979), 101.

Soviet Union, but there is no evidence that it played an important e in his alignment. The Ethiopian leader never manifested pro-Com- inist sympathies before seizing power, did not hesitate to purge ortho- x Communist groups ruthlessly, and bitterly resisted the establishment a Marxist-Leninist vanguard party. Thus, none of Mengistu's major cisions can be explained as a consequence of ideology.[41] Bandwagoning ay appear to have been a factor, with Mengistu seeking to appease wiet aggression. But Mengistu at no point attempted to appease or ign with his principal threat, Somalia, as bandwagoning would predict.

Only omnibalancing explains Mengistu's decision to align with the wiet Union, by recognizing that the United States, contrary to balance power assessments, was actually not providing adequate military sup- ort to the Ethiopian leadership. It supplied enough help for Ethiopia to al with the external threat posed by Somalia, but it did not provide iough assistance for Ethiopia to deal with the myriad of internal threats well. Foremost among these threats was the rebellion in the Ethiopian ovince of Eritrea, which involved far more Ethiopian troops than were itioned even along the Somali front.[42] Mengistu recognized that he uld not survive in power without satisfactorily coping with the Eri- ean revolt. Accordingly, gaining outside assistance to suppress the re- llion played a major role in his turn to the Soviet Union. The assump- on of balance of power that the need to resist internal threats, such as e one posed by Eritrea, is irrelevant in determining alignment is clearly rong in the case of Mengistu.[43]

Omnibalancing also recognizes that Mengistu's alignment to the So- et Union was not bandwagoning but actually a form of balancing be-

[41] There is a widespread consensus among analysts across the ideological spectrum that engistu and the Derg were not driven by ideological factors but, rather, acted simply to rvive in power. See, for example, Marina Ottaway and David Ottaway, *Ethiopia Empire Revolution* (New York: Africana, 1978), 149; Korn (fn. 39), 111; Fred Halliday and Maxine olyneux, *The Ethiopian Revolution* (London: Unwin, 1981), 99; Paul B. Henze, "Beyond e Ethiopian Famine: Anatomy of a Revolution II," *Encounter* (London) (July 1986), 15–27, 19; and Michael Chege, "The Revolution Betrayed: Ethiopia, 1974–1979," *Journal of Mod- African Studies* 17 (September 1979), 359–80. Mengistu's recent retreat from Marxism- ninism (which not so coincidentally coincides with reduced Soviet support) and his rees- blishment of military ties to Israel further support the view that ideological factors do not minate his decision making.

[42] Some useful works on Eritrea include Hagai Erlich, *The Struggle over Eritrea, 1962–1978* anford, Calif.: Hoover Institution, 1983); G. H. K. Trevaskis, *Eritrea: A Colony in Transi- n* (London: Oxford University Press, 1960); and Richard Sherman, *Eritrea: The Unfinished volution* (New York: Praeger, 1980). See also John Markakis, *National and Class Conflict the Horn of Africa* (London: Cambridge University Press, 1987), 104–45.

[43] Many analysts make the point that Mengistu could not survive without a satisfactory tlement of the Eritrean conflict. See, for example, Legum and Lee (fn. 39), 39; Suzanne sikas, *The Arc of Socialist Revolutions* (Cambridge, Mass.: Schenkman, 1982), 142; and assie (fn. 40), 37, 71.

havior not included in balance of power theory. By turning to the Soviet Union, Mengistu was not seeking to appease the principal threats arrayed against him (Ethiopia was far too unimportant to justify direct Soviet action). Rather, he was trying to resist them by appeasing the secondary threat posed by the Soviet Union. As the major supporter of Somalia and an indirect backer of the Eritrean rebellion, the Soviet Union clearly *contributed* to the threats that were most likely to topple Mengistu. But by backing the Somalis and the Eritreans, the USSR was also in the best position to undermine the threats they presented simply by withholding assistance and getting its allies to do the same. Balance of power may be correct that states will not normally appease threats because of fear of placing one's survival in the hands of an adversary. But, as demonstrated by Mengistu, leaders will appease secondary threats in order to defeat more pressing ones.

Anwar Sadat's alignment with the United States following the 1973 October War is another clear example of the limitations of balance of power theory.[44] After assuming power following Nasser's death in 1970, Sadat enhanced Egypt's alignment with the Soviet Union. A treaty of friendship and cooperation was concluded, and Moscow provided Egypt with record levels of military assistance. Soviet support reached a peak during the 1973 October War with Israel, enabling Sadat to achieve a smashing political victory. Nevertheless, after the war Sadat decided to forsake his alignment with Moscow and turn to the United States—the backer of his chief external threat—for support.

As in the case of Mengistu, neither ideology nor bandwagoning explains Sadat's alignment with the United States. Ideological explanations for Sadat's realignment decision—that he underwent a fundamental change in his world view—are not very convincing. At the very least, that Sadat could profess being comfortable with the Soviet Union and then the United States demonstrates an ideological commitment too

[44] Some good background treatments of Anwar Sadat's Egypt, including his decision to reject his alignment with the Soviet Union and turn to the United States, include Anwar el-Sadat, *In Search of Identity: An Autobiography* (New York: Harper and Row, 1978); Raymond William Baker, *Egypt's Uncertain Revolution under Nasser and Sadat* (Cambridge: Harvard University Press, 1978); Raymond A. Hinnebusch, Jr., *Egyptian Politics under Sadat* (Cambridge: Cambridge University Press, 1985); A. I. Dawaisha, *Egypt in the Arab World: The Elements of Foreign Policy* (New York: Macmillan, 1976); Mohammed Heikal, *The Sphinx and the Commissar: The Rise and Fall of Soviet Influence in the Arab World* (London: Collins, 1978); Shimon Shamir, "Egypt's Reorientation towards the U.S.: Factors and Conditions of Decision Making," in Haim Shaked and Itamar Rabinovich, eds., *The Middle East and the United States: Perceptions and Policies* (New Brunswick, N.J.: Transaction, 1980); Mark Cooper, *The Transformation of Egypt* (Baltimore, Md.: Johns Hopkins University Press, 1982); Alvin Z. Rubinstein, *Red Star on the Nile: The Soviet-Egyptian Influence Relationship since the June War* (Princeton: Princeton University Press, 1977); and Robert O. Freedman, *Soviet Policy toward the Middle East since 1970* (New York: Praeger, 1975).

weak to dictate the choice of alignment partners. Similarly, bandwagoning was not in evidence, as Sadat made no effort to appease or align with Israel.

Only omnibalancing's emphasis on internal threats explains Sadat's decision to align with the United States. Omnibalancing recognizes that Israel, the primary external threat faced by Egypt, did not directly threaten Sadat's hold on power. Since the end of the War of Attrition in 1970 (in which Israeli and Egyptian forces exchanged fire across the Suez Canal), Israel had had no offensive designs against Sadat's Egypt. The Israeli threat stemmed not from Israel itself but from the effect of the continuing Israeli occupation of Egyptian land (a result of the 1967 war) on various Egyptian groups. The Egyptian military, workers, intellectuals, and students had grown increasingly impatient with the continuing stalemate and sought to force the Egyptian leader to regain the territories from Israel. For Sadat to deal with the threats from the Egyptian groups, he had to force Israel to return the lands it had seized. The Soviet Union, however, fearing a confrontation with the United States and believing that Egypt would lose any war with Israel, would not provide Sadat with the capability to end the Israeli occupation. Moscow's unwillingness to do what was necessary to extricate Sadat from the continuing stalemate exacerbated the domestic threats he confronted and eventually led him to dissolve the Egyptian-Soviet alignment.[45]

Sadat subsequently aligned with the United States out of a recognition that that was the best means of balancing against his principal, domestic threats. The United States, by virtue of its support for Israel, was the only great power willing and able to force an end to the Israeli occupation of the Sinai. Thus, the reasons for Sadat's alignment with the United States only become clear when assessed in light of the domestic threats he faced and the impact of the Israeli occupation on those threats.

Both cases clearly show how omnibalancing offers the necessary correctives to balance of power to explain Third World alignments. Mengistu and Sadat both appeased an international threat (posed by the Soviet Union and the United States, respectively) in order to defeat a more pressing domestic challenge. The notion that Third World leaders appease (or bandwagon with) an external threat in order to balance against a domestic threat is not as contradictory as it may appear. As Robert Jervis argues, politics in a complex environment results in unintended consequences. Instead of attempting to achieve a goal by heading straight for it, one is often better off adopting an indirect approach or even one

[45] Extensive accounts of the internal threats besetting Sadat can be found in Baker (fn. 44); Hinnebusch (fn. 44); and Cooper (fn. 44).

that heads in the opposite direction.[46] As seen by Mengistu and Sadat—Third World leaders confronting a great number of threats from many different sources—a simple balance of power approach would not have maximized the chances of survival. Instead, a Third World leadership often stands the best chance of defeating the threats it confronts by adopting a policy that is indirect and seemingly opposed to what it wishes to achieve.

The examples of Mengistu and Sadat also show how focusing on the interests of the leadership elite better explains alignment than does focusing on the interests of the state as a whole. It is unclear whether Mengistu's alignment with the Soviet Union or Sadat's alignment with the United States benefited their respective countries. Ethiopia, in particular, suffered as a result of its alignment with the Soviet Union.[47] But if the alignments did not benefit Ethiopia or possibly even Egypt, they unquestionably benefited the leaders of those counties by enhancing their hold on power. In comparison with other leaders, those in the Third World tend to face more threats and pay a higher price for failure; presumably, then, the concern for political survival that characterized Mengistu's and Sadat's alignment decisions exists throughout the Third World.

The two cases furthermore call into question the contention that Third World leaders will align with the outside state that provides them with the most benefits. Both leaders rejected patrons who gave them record assistance for others whose level of support could not be guaranteed. Mengistu had little reason to believe that Soviet economic assistance would ever approach the quantity and quality of that provided by the United States, and Sadat could not reply on a United States beholden to Israel to match the amount of weaponry supplied to Egypt by the Soviet Union. Both leaders did have reason to believe, however, that realigning would place them in a better position to defeat the principal threats confronting them. In deciding to change patrons, therefore, they behaved more consistently with omnibalancing's emphasis on aligning to resist threats rather than on aligning with the better provider.

Finally, both cases demonstrate how focusing exclusively on external threats—and ignoring internal threats—can fail to explain or predict Third World alignments. Internal threats, either alone or backed by outside states, played a central role in the alignments of Ethiopia and Egypt.

[46] Jervis, "Systems Theories and Diplomatic History," in Paul Gordon Lauren, ed., *Diplomacy: New Approaches in History, Theory and Policy* (New York: Free Press, 1979), 218.

[47] For how aligning with the USSR exacerbated Ethiopia's problems, see Paul B. Henze, "Communism and Ethiopia," *Problems of Communism* 30 (May–June 1981), 55–74.

Such threats continue to determine Third World alignments today. In Latin America, for example, requests for American security assistance have come from the leaders of Colombia and Bolivia (which face threats from well-armed drug dealers). In the Middle East domestic pressure from Palestinians has contributed to the decision on the part of Jordan's King Hussein to align his state more closely with Iraq, and in Southeast Asia, Cambodia's ultimate alignment will be determined by that country's civil war. The determinative role of internal threats in Third World alignments should not be surprising given the prevalence of such threats. What is surprising is the belief that balance of power can be applied generally to explain Third World alignment, despite its neglect of the kind of threats that are most likely to confront the leaders of the Third World.

The Defense of Balance of Power

Supporters of balance of power have responded to the prevalence of internal threats in the Third World generally and to their impact on specific countries by maintaining that it does not invalidate the balance of power theory. Some adherents of the theory, such as Waltz, recognize that violence and the use of force to deal with it occur as often (or more so) within states as between states. For Waltz, this means that neither the occurrence of violence nor the use of force per se can be used as a standard by which to distinguish domestic from international politics. Instead, he argues, the distinction is marked by government's monopoly on the *legitimate* use of force to deal with violence. Citizens, therefore, need not worry about protecting themselves; they can appeal to the government for assistance. In international politics, by contrast, states can only rely on themselves for defense. Thus, international politics is a system of self-help, whereas domestic politics is not.[48]

In a consideration of Third World alignment, however, this largely cogent analysis is problematic. First, the notion of a legitimate government protecting its citizens from private acts of violence often does not apply: for large groups of people in the Third World the government is neither legitimate nor a protector. Whether one is a southern Sudanese seeking autonomy in Nimeiri's Sudan, or an Eritrean seeking independence in Mengistu's Ethiopia, or an individual seeking just to survive in Pol Pot's Cambodia, the government is less of a guardian against threats than it is the chief threat. Second, Waltz's observations are largely irrel-

[48] Waltz (fn. 3), 103-4.

evant in terms of the alignment decisions of Third World leaders. Even if it is granted that a government has a monopoly on the legitimate use of force, its leadership must still act to guard itself from all threats, including "illegitimate" ones and those that originate within their own countries. A Third World leadership struggling to survive bases its alignment decisions on how best to protect itself from the threats it faces, not on abstract considerations of whether the threat is "legitimate" or whether its source is domestic or international.

It has also been argued that it does not matter whether the assumptions of balance of power are true as long as the theory based on those assumptions accurately predicts what occurs. Assumptions are seen as simplifications of the truth, their role being not to describe what actually happens but to enable theories to be constructed. According to Waltz, "We can freely admit, then, that states are in fact not unitary, purposive actors." As such, the fact that Third World states do not conform to the ideal assumed by balance of power theory is irrelevant. So long as balance of power theory accurately predicts the behavior of Third World states, the thinking continues, it remains a valuable theory for the Third World.[49]

As this analysis has shown, however, it has not yet been demonstrated that Third World states do in fact conduct themselves in a manner predicted by the balance of power theory. Moreover, even if balance of power theory accurately predicts the behavior of most Third World states, it is worth knowing whether it is right for the wrong reasons. Especially where theory is used to guide policy (as balance of power does), it is critical to know which policies are likely to produce which outcomes. Basing one's policies on overall outcomes might prove catastrophic in a specific, vital case.

Most important, the pursuit of knowledge is worthy in its own right. If assumptions do not conform to reality, regardless of the value of the theories they underlie, they should be revised. This is especially the case when those assumptions grew out of the study of states the present nature and historical experiences of which are vastly different from what one finds in the Third World. At the very least, refining our assumptions in light of Third World realities holds out the prospect of an improved theory. This is particularly important for theories such as balance of power, whose explanatory and predictive value leaves much room for improvement.

Finally, some have argued that although many states in the Third

[49] Waltz (fn. 5), 37–38.

World do not follow the dictates of balance of power when aligning, that does not detract from the value of the balance of power theory. As Waltz argues (in the context of defending the primacy of states over transnational actors), "So long as the major states are the major actors, the structure of international politics is defined in terms of them."[50] Since the major states are not Third World states, their behavior is apparently not sufficient to invalidate theories of international politics. Moreover, those states that are driven by internal threats (and thus do not follow balance of power dictates) are likely to be weak, in which case they will not affect the global balance of power anyway.[51]

Why then concern ourselves with these exceptions? First, it matters little that international politics may be defined by the major states if this definition does not help in the understanding of why Third World leaders make the alignment decisions they do. Second, any theory of international relations that is of questionable applicability for most of the world's states and people requires refinement. Finally, as discussed below, understanding the Third World is central to the understanding of some of the fundamental issues in international relations.

The Significance of the Third World

The importance of the Third World is most clearly seen when considering questions of war and peace. Understanding why states go to war has traditionally been at the heart of the study of international relations. Since World War II wars between and within non–Third World states have virtually ceased to exist. Many reasons are offered to explain this unprecedented phenomenon, including fear of nuclear war, economic entanglement, desire to avoid another World War II, and a growing acceptance of Western liberal democracy.[52] The disagreements about why war has ended outside the Third World should not, however, obscure the recognition that conflict continues unabated in the Third World. Such conflict persists in threatening both Third World and non–

[50] Waltz (fn. 3), 94.

[51] Walt makes this point in connection with bandwagoning; see Walt (fn. 2), 179.

[52] On the possible reasons for the absence of war among states outside the Third World, see Michael W. Doyle, "Kant, Liberal Legacies, and Foreign Affairs," *Philosophy and Public Affairs* 12 (Summer–Fall 1983), 205–35, 323–53; idem, "Liberalism and World Politics," *American Political Science Review* 80 (December 1986), 1151–69; Bruce Russett, "The Politics of an Alternative Security System: Toward a More Democratic and Therefore More Peaceful World," in Burns Weston, ed., *Alternatives to Nuclear Deterrence* (Boulder, Colo.: Westview Press, 1989); Gaddis (fn. 18), esp. chap. 8; John Mueller, "The Essential Irrelevance of Nuclear Weapons: Stability in the Postwar World," *International Security* 13 (Fall 1988), 55–79; and Francis Fukuyama, "The End of History," *National Interest* 16 (Summer 1989), 3–18.

Third World states. Because war has become a virtual Third World monopoly, it is impossible to understand the causes and possible solutions of international conflict without understanding the Third World.

Present trends point to the increasing importance of the Third World in matters of international security. The proliferation of weapons and weapons-related technology has dramatically enhanced the strength of many Third World states. Approximately a dozen Third World countries have chemical or biological weapons, and roughly the same number maintain ballistic missiles.[53] Several Third World countries have or are very close to having nuclear weapons. Even without exotic weapons, Third World militaries have become formidable institutions. Many Third World countries (for example, Iraq, India, and Vietnam) have military forces larger and better equipped than those of most European countries. A dozen deploy a thousand or more battle tanks. In most cases, the increasing strength of Third World armies has not been accompanied by a commensurate increase in stability.

Ironically, the apparent lessened interest of the Soviet Union in the Third World, if continued, will enhance the Third World's role in issues of international security. A Soviet departure from the Third World would not end the threats to global security posed by Third World states. On the contrary, for some threats (for example, nuclear proliferation, regional conflict), a Soviet withdrawal could eliminate an important constraint. Most important, were the Soviets to abandon their efforts to expand their influence throughout the world, the Third World would be left as the principal threat to international stability. At the very least, therefore, the *relative* importance of the Third World to international security concerns would increase in the wake of overall Soviet retrenchment.

Many of the issues that are likely to dominate international relations for the foreseeable future have a critical Third World component. Economically, the exploding growth of the newly industrialized countries of East Asia and the steady development of major Third World powers such as India and Brazil will make them powerful forces in determining the shape and direction of the international economy.[54] And as the

[53] For a list of Third World countries with chemical arms, see *Newsweek*, September 19, 1988; for lists of Third World countries with ballistic missiles, see W. Seth Carus, "Missiles in the Middle East: A New Threat to Stability," Washington Institute for Near East Policy, *Policy Focus* 6 (June 1988), 1–9; and Aaron Karp, "Ballistic Missiles in the Third World," *International Security* 9 (Winter 1984–85), 166–95. On nuclear proliferation, see Leonard S. Spector, with Jacqueline R. Smith, *Nuclear Ambitions* (Boulder, Colo.: Westview Press, 1990).

[54] For more on the growing economic power of the Third World, see U.S. Department of Defense, *Sources of Change in the Future Security Environment* (Washington, D.C.: G.P.O. 1988), 4.

world's dependence on petroleum increases in the coming years, the importance of the Persian Gulf will be enhanced.[55] The Third World also plays a major role in the global environment. There is increasing evidence that the burning of tropical forests contributes to the greenhouse effect, whereby the climate of the earth is substantially warmed. Dealing with this problem will require the cooperation of the Third World states in which the forests are located. In addition, problems such as overpopulation and drug trafficking are increasingly dependent on the Third World for resolution. Far from being marginal actors, the Third World is central to the course of international relations, and it will become even more so.

Omnibalancing and the Third World

Omnibalancing's explanation of why Third World leaders align as they do addresses the implications of the increasing importance of the Third World in two critical ways. First, understanding why Third World leaders make the alignment decisions they do enables other states to influence those decisions in order to encourage favorable alignments. Omnibalancing's emphasis on the need to meet the security concerns of Third World leaders in order to secure an alignment with them suggests approaches (for example, a greater emphasis on dealing with internal threats) that differ from those deriving from the focus of balance of power on protecting states from other states. If omnibalancing provides a better guide for policy than does balance of power, states that follow its precepts are more likely to secure the alignments they seek. And states that secure alignments with Third World leaders, at the very least, insulate themselves from the growing power and threats those leaders can pose.

The emphasis on the overriding need of Third World leaders to act in ways that will ensure their hold on power also has implications for a broad range of Third World behavior that goes beyond decisions on alignment. As seen in the cases of Egypt and Ethiopia, powerful forces are at work when Third World leaders take the risky and problematic step of breaking an alignment with one great power in order to align with another. This suggests that other decisions relating to both foreign and domestic policy should also be examined in terms of the Third World leader's drive to remain in power. Insights into questions concerning when Third World leaders will go to war or how to enlist their cooperation in solving global problems could well be gained by focusing

[55] United States Department of Energy, *Energy Security: A Report to the President of the United States* (Washington, D.C.: G.P.O. 1987).

on the need of Third World leaders to ensure their survival in the face of both external and internal threats.[56]

The Third World is and will continue to be central to the study of international relations. For both theoretical and pragmatic reasons, understanding Third World foreign policies, including why Third World states align as they do, is essential for the broader understanding of world politics. In seeking such an understanding, it makes little sense to rely on a theory derived from a different historical experience that fails to address the reality of the Third World. That reality is dominated by the efforts of Third World leaders to do what they can, including seeking outside support, to survive in power. Ignorance of this simple truth prevents an understanding of the politics that determines the destiny of most of the world.

[56] For example, Richard Ned Lebow cites the political vulnerability of a leader as one of the key factors that can induce a policy of brinkmanship; Lebow, ***Between Peace and War: The Nature of International Crisis*** (Baltimore, Md.: Johns Hopkins University Press, 1981), 69–79.

Review Articles

THE SECURITY PROBLEMATIC OF THE THIRD WORLD

By MOHAMMED AYOOB

Edward E. Azar and Chung-in Moon, eds., *National Security in the Third World: The Management of Internal and External Threats.* College Park, Md.: Center for International Development and Conflict Management, University of Maryland, 1988, 308 pp.
Nicole Ball, *Security and Economy in the Third World.* Princeton: Princeton University Press, 1988, 432 pp.
Robert S. Litwak and Samuel F. Wells, Jr., eds., *Superpower Competition and Security in the Third World.* Cambridge, Mass.: Ballinger, 1988, 295 pp.
Caroline Thomas, *In Search of Security: The Third World in International Relations.* Boulder, Colo.: Lynne Rienner, 1987, 228 pp.

I

TWO major events have shaped the political contours of the postwar world. The first is the awesome destructive capability of nuclear weaponry, which, as institutionalized in the doctrine of mutual assured destruction (MAD), has prevented the outbreak of major war between the two dominant powers in the international system and, until recently, had frozen the strategic situation in Europe in a bipolar mold. The second is the entrance of unprecedented numbers of new members into the system of states as a result of the decolonization process—such that the newcomers now constitute a majority among the membership of that system. Although the latter event has had as far-reaching effects as the former on the workings of the international system, it has unfortunately not received attention in the literature on international relations commensurate with its actual and potential impact on international affairs.

Despite the moral dilemma related to the capacity of nuclear weapons for mass destruction, the existence of superpower nuclear arsenals with second strike capacities helped during the last four and a half decades to stabilize the global balance of power and make it relatively immune to transient shifts in the capabilities of the great powers. By contrast, the influx of the weak, intruder majority of Third World states into the

international system[1] introduced a great deal of fluidity, and therefore instability, into that system. In strategic terms, this resulted primarily from the fact that there existed a large group of "floating" states with no alliance commitments, a group that was, in a sense, "up for grabs" by the highest bidder or the great power with the largest capability to help or harm it. Consequently, Third World regions became gray areas of the globe to which cold war energies, frustrated in Europe by the existence of MAD, were diverted. They became the primary site of the new "great game" played out by the United States and the Soviet Union.

As a result of the superpowers' involvement in the Third World, particularly in conflictual situations, a great deal of the international relations literature on the Third World has been written from the perspective of superpower competition for power and influence in the strategic regions of the developing world. And it followed that the security of Third World states and regions has been analyzed primarily from the point of view of American and/or Soviet interests and concerns.[2] Other than individual country studies, little has been written in a systematic fashion about the interaction of Third World states with the international system,[3] in particular, about their overriding concern with security in terms of reducing the vulnerabilities of their structures, institutions, and regimes.[4]

The books under review attempt to fill this gap in the existing literature on the subject. The volumes by Azar and Moon and by Thomas address the issue of the interaction of the Third World states with the international system in the light of these states' demonstrated weaknesses and vulnerabilities. Both books emphasize the search for security on the part of these states as the major determinant of their external and internal behavior. In their own contribution, Azar and Moon (chap. 4) em-

[1] For a discussion of this phenomenon, see Mohammed Ayoob, "The Third World in the System of States: Acute Schizophrenia or Growing Pains?" *International Studies Quarterly* 33 (March 1989), 67–79.

[2] For example, see Michael Nacht, "Toward an American Conception of Regional Security," *Daedalus* 110 (Winter 1981), 1–22; and S. Neil MacFarlane, "The Soviet Conception of Regional Security," *World Politics* 37 (April 1985), 295–316.

[3] There are, however, significant exceptions to this rule and they include Robert I. Rothstein, *The Weak in the World of the Strong: The Developing Countries in the International System* (New York: Columbia University Press, 1977); Stephen D. Krasner, *Structural Conflict: The Third World against Global Liberalism* (Berkeley: University of California Press, 1985); Robert A. Mortimer, *The Third World Coalition in International Politics*, 2d ed. (Boulder, Colo.: Westview Press, 1984); and Jacqueline A. Braveboy-Wagner, *Interpreting the Third World: Politics, Economics, and Social Issues* (New York: Praeger, 1986).

[4] For a commendable effort at bringing together analyses of the security problems and policies of important Third World countries within some sort of a common framework, see Edward A. Kolodziej and Robert E. Harkavy, eds., *Security Policies of Developing Countries* (Lexington, Mass.: Lexington Books, 1982). See also Abdul-Monem M. Al-Mashat, *National Security in the Third World* (Boulder, Colo.: Westview Press, 1985).

phasize what they call the "software" side of the security problematic in the Third World as opposed to the traditional Western analyses of security, which tend to concentrate on the "hardware" side of the problem. They operationalize the concept of "security software" by disaggregating it into three primary components—legitimacy, integration, and policy capacity—and attempting, in the light of these variables, to formulate a typology of Third World states based upon eight clusters. These in turn are dependent upon the different ways in which the three dimensions of state security, namely, threats (security environment), hardware (capabilities), and software, interact in the case of particular states.

Thomas also attempts to broaden the scope of the discussion of Third World security but in directions somewhat different from those chosen by Azar and Moon. In her own words:

> A basic theme running through the book is that security in the context of the Third World states does not simply refer to the military dimension, as is often assumed in Western discussions of the concept, but to the whole range of dimensions of a state's existence which are already taken care of in the more-developed states, especially those of the West. . . . [F]or example, the search for the internal security of the state through nation-building, the search for secure systems of food, health, money and trade, as well as the search for security through nuclear weapons. (p. 1)

This all-inclusive definition of security, while valuable as an antidote to the traditional military-oriented definition of the term, nevertheless runs the risk of making the concept so elastic as to detract seriously from its utility as an analytical tool.

For this reason it is preferable to define security in relation to vulnerabilities that threaten, or have the potential, to bring down or significantly weaken state structures, both territorial and institutional, as well as the regimes that preside over these structures and profess to represent them internationally. According to this definition, the more a state or regime falls toward the invulnerable end of the vulnerable-invulnerable continuum, the more secure it is. This is a definition that assumes the basic primacy of political variables in determining the degree of security that states and regimes enjoy. Different types of vulnerability, including those of the economic and ecological varieties, become integral components of this definition of security only if and when they become acute enough to take on overtly political dimensions and threaten state boundaries, state institutions, or regime survival. In other words, debt burdens, or even famines, do not become part of the security calculus for the purpose of this definition unless they threaten to have political outcomes that affect the survivability of states (in either the territorial or the institutional sense or both) or of governing elites within those states.

The volume edited by Litwak and Wells is in its orientation the most traditional of the four books under review, in that it emphasizes the East-West dimension of security problems in the Third World. Nevertheless, it also acknowledges the "turbulent nature of the target environment itself" and the fact that "regional conflict[s] create the preconditions for outside power intervention and involvement" (p. xi). Furthermore, the editors do not hesitate to criticize the superpowers, who, they charge, "proclaim their sensitivities to the nuances of regional politics" but tend "to view Third World developments as a function of East-West competition" (p. xii). Overall, the volume attempts to juxtapose the geopolitical (East-West) and regionalist approaches to the study of security in the Third World. The two-part organization of the book reflects this approach. The first part, consisting of four chapters, is devoted to the perceptions and policies of the developed states, and the second, comprising seven chapters, deals with issues of superpower competition in the Third World on a region-by-region basis.

The fourth and last book under review, Ball's study of security and economy in the Third World, takes a rather different approach by trying to explore the relationship between what appear to be the two most important concerns of Third World policymakers—security and development. Ball poses and attempts to answer the query: "Does expenditure in the security sector of Third World countries hinder their development, or does it, as some analysts have suggested, promote the development process?" (p. xiii). More specifically, according to Ball, the volume "seeks to incorporate the security sector into the debate on the development process in the Third World" (p. xxv). This very important, though relatively neglected, field of inquiry has a great bearing upon the question of how Third World governments perceive and prioritize their objectives and in what fashion they allocate resources to the two sectors that impinge directly on the legitimacy of both states and regimes in the Third World.

Together the four volumes—by attempting to broaden the definition of security, by opening up relatively new and neglected areas of enquiry, and by pointing toward fresh directions for research—raise some very important issues in the field of Third World security studies. These are issues that, in turn, need to be addressed from both historical and comparative perspectives, as follows: (1) How does the concept of security as applied to the Third World context differ from its traditional use in the international relations literature? (2) What are the factors that inhere within Third World states that can help explain this difference? (3) In what ways does the interaction of Third World states with the interna-

tional system affect the security of the former? (4) Are there specific factors related to technology in the late twentieth century that affect the security of Third World states in ways that are unique to the developing countries? (5) What is the relationship between the security and developmental concerns of Third World states, and how does the interaction between these two preoccupations of Third World state elites affect the levels of legitimacy enjoyed by Third World states and regimes?

II

The term *security* as it has been traditionally used in international relations literature is based on two major assumptions: one, that threats to a state's security principally arise from outside its borders, and two, that these threats are primarily, if not exclusively, military in nature and usually need a military response if the security of the target state is to be preserved. These assumptions were best summed up in Walter Lippmann's celebrated statement that "a nation is secure to the extent to which it is not in danger of having to sacrifice core values, if it wishes to avoid war, and is able, if challenged, to maintain them by victory in such a war."[5] Lippmann's definition, according to Arnold Wolfers, "implies that security rises and falls with the ability of a nation to deter an attack, or to defeat it. This is in accord with the common usage of the term."[6]

Even those scholars who have differed from this starkly state-centered realist perspective and focused on *international* rather than *national* security have been primarily concerned with reconciling national security (in terms of reducing external threats to the security of a state, especially of a major power) with systemic security concerns. They have taken their philosophical cue from authors like Martin Wight and Hedley Bull, who have argued, to quote Wight, that

> if there is an international society, then there is an order of some kind to be maintained, or even developed. It is not fallacious to speak of a collective interest, and security acquires a broad meaning: it can be enjoyed or pursued in common.[7]

Indeed, the earliest of the twentieth-century proponents of international security—the "idealists" of the first three decades[8]—refused to distin-

[5] Lippmann, *U.S. Foreign Policy: Shield of the Republic* (Boston: Little Brown, 1943), 51.

[6] Wolfers, *Discord and Collaboration: Essays on International Politics* (Baltimore, Md.: Johns Hopkins University Press, 1962), 150.

[7] Wight, "Western Values in International Relations," in Herbert Butterfield and Martin Wight, eds., *Diplomatic Investigations* (London: Allen and Unwin, 1966), 103.

[8] For a representative sample of idealist thought, see Norman Angell, *The Great Illusion*, 4th ed. (1909; New York: Putnam's, 1913).

guish the security of the parts from that of the system as a whole. The post-Second World War breed of system-centered scholars has been more discriminating than its predecessors. They have argued from the assumption that the various segments of the international system are interlinked to such an extent that their security and welfare are dependent upon each other. While much of the initial impetus for this line of argument came from the awesome concentration of nuclear weaponry in the hands of the two superpowers and the periodic crises in their relations from the Berlin blockade of 1948 to the Cuban missile crisis of 1962, the economic problems that the leading Western industrialized states faced from the early 1970s, including the two oil shocks of 1973–74 and 1978–79, led to the crystallization of the "interdependence" argument.[9]

What is most interesting for our purpose is that both these dominant strands of security thinking (in their many variations) defined the concept of security in external or outward-directed terms, that is, as external to the commonly accepted unit of analysis in international relations: the state. This definition and the process by which it was reached were understandable because both reflected a particular trajectory of historical development that could be traced back at least to the Peace of Westphalia if not earlier. Between 1648 (to use it as a symbolic date) and 1945 the evolution of the European system of states and its interaction with the domestic political processes of state building and national consolidation within the major European powers led to the legitimation both of the system and of the individual participants (at least of those twenty-five or so that survived the processes of war and change in the European system and emerged as modern states by 1900). These two trends—of interaction among sovereign states and of greater identification of individuals with their respective states—strengthened each other and in doing so firmly laid the foundations of the intellectual tradition in which, at least in terms of the literature on diplomatic history and international relations, security became synonymous with the protection of a state's vital interests and core values from external threats.

Developments since 1945 strengthened the traditional Western notions about security. In dividing the Western world (that is, Europe and its offshoots) into two halves and in stabilizing that division until recently by means of a mutual balance of terror, the cold war (and its later manifestation, detente) froze the predominant Western connotation of security in a bipolar mold. The concept of alliance security was, therefore, superimposed on the concept of state security, while its essential, exter-

[9] For example, see Robert O. Keohane and Joseph S. Nye, *Power and Interdependence* (Boston: Little Brown, 1977).

nally directed thrust remained unchanged. Moreover, by making the security of major industrial states of Europe and North America the central concern of the security of the international system as a whole, the dominant strand in Western strategic thinking increasingly obliterated even the distinction between the realist (state-centric) and idealist (system-centric) approaches to the study of international security.

The application of this historically conditioned definition of the concept of security to the analysis of Third World situations has, however, created major conceptual problems. This is so because the three major characteristics of the concept of state security as developed in the Western literature on international relations— namely, its external orientation, its strong linkage with systemic security, and its binding ties with the security of the two major alliance blocs—are, if not totally absent, at lest thoroughly diluted in the Third World. Thus, the explanatory power of the concept, as traditionally defined, is vastly reduced when applied to Third World contexts.

The first and, in a sense, the fundamental attribute of the Western concept of security (in that it is a corollary of the doctrine of state sovereignty in its pure and pristine form) is external directedness. But it is clear that in the Third World, despite the rhetoric of many of its leaders, the sense of insecurity from which states suffer emanates to a substantial degree from within their boundaries rather than from outside. This is borne out by, among other studies, the findings of a recent project on the security perceptions of leaders of Southeast Asian states. That study presents the conclusion that "most Southeast Asian leaderships, like their counterparts in the rest of the Third World, are preoccupied primarily with internal threats to the security of their state structures and to the regimes themselves."[10] While this does not mean that external threats are nonexistent, it does imply that where external threats do exist they often attain saliency primarily because of the insecurities and conflicts that abound within Third World states. Furthermore, it can be argued that these internal conflicts and insecurities frequently get transformed into interstate conflicts because of their spillover effects into neighboring states that often suffer from similar domestic insecurities. Several contributions to the Azar-Moon volume, particularly those by Barry Buzan (chap. 2) and by the editors (chaps. 1 and 4), as well as the second chapter of Thomas's book, which deals with nation building and the search for security, highlight this internal dimension of the Third World states'

[10] Mohammed Ayoob and Chai-Anan Samudavanija, "Leadership and Security in Southeast Asia: Exploring General Propositions," in Ayoob and Samudavanija, eds., *Leadership Perceptions and National Security: The Southeast Asian Experience* (Singapore: Institute of Southeast Asian Studies, 1989), 256.

security problems and its capacity to become enmeshed in and, not infrequently, to generate interstate conflict.

The Third World's weak linkage with the systemic security agenda further circumscribes the utility of the traditional concept of security in explaining the problem (or problems) of security that Third World states face. This reflects the remarkable difference between the respective relationships of the security concerns of the Third World states, on the one hand, and those of the developed countries, on the other, to the security and stability of the international system as a whole. The Third World's relative unimportance to the central strategic balance is, paradoxically, borne out by the fact that

> during the postwar era, the Third World has been a principal arena of East-West rivalry. From Southeast Asia to the Middle East to Southern Africa to Central America, the superpowers have found themselves on opposing sides of regional conflicts, locked in a global competition for influence. (Litwak and Wells, ix)

The very fact that the superpowers chose the Third World as the arena in which they could afford to be "locked in a global competition for influence" in the thermonuclear age demonstrates the low priority they attached to gains and losses in the Third World and the vast distance that separated their Third World concerns from their vital interests, which were, and are, protected by the nuclear balance of terror. It is no wonder then that conflicts have proliferated in the Third World, while the industrial and strategic heartland of the globe has been free of major interstate conflict since the end of the Second World War. Systemic security has therefore often contributed to insecurity in the Third World.

The close linkage between alliance security and state security that has been such a prominent feature of the postwar political landscape in Europe has been conspicuous by its absence in the Third World. While several Third World states have been allied with one or the other superpower, such alliances have been either fluid and temporary (as in the case of Egypt and Somalia) or inadequate deterrents to regional conflicts involving superpower allies (for example, Vietnam and Iraq) or incapable of preventing the dismemberment of at least one aligned state (Pakistan). The nature of alliances and of superpower commitments to their allies in the Third World are therefore vastly different from the character of alliances and of alliance commitments in the developed world. Alliance security, in contrast to the postwar situation in Europe, is not synonymous with, or even inextricably tied to, the security of even the most overtly aligned states in the Third World.[11]

[11] Israel is the only exception to this rule because of the intensity of one superpower's

For all the three dimensions of the traditional definition of security—its external orientation, links with systemic security, and the correspondence with alliance security—the situation in the Third World is radically different from that prevailing at the heart of the global strategic system, which includes the two superpowers, Europe and its offshoots, and Japan. The security of Third World states therefore needs to be looked at from a perspective that differs somewhat from the one that is prevalent in the Western literature on international relations.

III

This leads us to our next question: Are there any factors that inhere within Third World states that can help explain this difference in the Third World state's security problematic as compared with the paradigm of security that is dominant in the international relations and strategic studies literature?[12] The principal problem that seems to distort a great deal of Western analysis of the security of Third World states is the tendency to compare states (that is, industrialized states with developing ones) that are unlike each other in many respects. This is especially so in relation to the crucial variable of state making, where the commonality is simply that both are in formal possession of juridical statehood. This, however, does not preclude the possibility of Third World states eventually approximating more closely the ideal type of the modern industrialized state (which is the reference point of most security analysts), given adequate time to complete the prerequisite twin processes of state making and nation building.

Time is, therefore, the crucial variable in explaining the difference in the security concerns of the two sets of states. Most security analysts tend to gloss over the fact that today's modern states—which are internally relatively cohesive, possess rational bureaucratic structures as well as a good deal of "infrastructural power,"[13] and are responsible to their peo-

commitment to its security, as defined largely by Israel itself. This, in turn, is related to the fact that Israel is a domestic political issue in the United States and not merely a foreign policy concern. Moreover, Israel, in terms of its ideological origins, the organization of its society and polity, the composition of its elite, and its links with strong and important European and American constituencies is not a Third World state. In other words, Israel may be physically located *in* the Third World but, in terms of the defining characteristics of the Israeli state, is not *of* the Third World. For details of the nature and evolution of the special relationship between the United States and Israel, see Nadav Safran, *Israel: The Embattled Ally* (Cambridge, Mass.: Belknap Press, 1981).

[12] For a critique of the application (or misapplication) of the strategic studies paradigm to the Third World, see Bahgat Korany, "Strategic Studies and the Third World: A Critical Evaluation," *International Social Science Journal*, no. 110 (1986), 547–62.

[13] Michael Mann has used the term *infrastructural power* to denote "the capacity of the state actually to penetrate civil society, and to implement logistically political decisions throughout

ple as well as responsive to the demands of their populations—were not created overnight. They went through a long period of gestation (during which most embryonic and also some not-so-embryonic states were aborted) before they acquired the functional capacities as well as the legitimacy they have today in the eyes of the populace that they encompass territorially and over which they preside institutionally.

It is worth noting in this context the testimony of two leading scholars of state making in Europe. According to Joseph Strayer:

> While the sovereign state of 1300 was stronger than any competing political form, it was still not very strong. ... It took four to five centuries for European states to remedy their administrative deficiencies, and to bring lukewarm loyalty to the white heat of nationalism.[14]

Charles Tilly makes the same point even more forcefully:

> The seventeenth and eighteenth century focus [of his edited volume] has us dealing with periods in which, for most of Europe, both the primacy and the ultimate form of the state were much in doubt. Perhaps that is the most important historical insight the book has to offer: as seen from 1600 or so, the development of the state was very contingent; many aspiring states crumpled and fell along the way.[15]

Most European political entities had to endure the precarious balance between success and failure for centuries before their statehood was assured; during that time their state makers were constantly preoccupied with the problem of consolidating their power and control *within* the territories they aspired to dominate. Seen in light of the European historical experience, then, the magnitude of the internal security problems faced by the new states of the Third World today is not all that astounding. These problems assume inflated dimensions only when compared with the "finished" products in Western Europe and North America. And, indeed, some of those, despite the centuries available to them, have yet to establish their unconditional legitimacy with some, and sometimes significant, segments of their populations. (Witness Northern Ireland, Quebec, and the Basque country, to mention only a few.)

Recent events have clearly demonstrated that the Soviet Union, despite its claim to be the successor state to tsarist Russia with its long history of statehood, does not even come close to the model of the cohe-

---

the realm"; Mann, "The Autonomous Power of the State: Its Origins, Mechanisms and Results," in John A. Hall, ed., *States in History* (Oxford: Basil Blackwell, 1986), 113.

[14] Strayer, *On the Medieval Origins of the Modern State* (Princeton: Princeton University Press, 1970), 23, 57.

[15] Tilly, "Reflections on the History of European State-Making," in Tilly, ed., *The Formation of National States in Western Europe* (Princeton: Princeton University Press, 1975), 7.

sive nation-state or even of a multinational federation whose institutional and territorial legitimacy is accepted by the overwhelming majority of its diverse ethnic and national groups. The same applies to the states in the Balkans, created out of the debris of the Habsburg and Ottoman empires, which generally fall midway between the model of the relatively cohesive nation-states of Western Europe and the postcolonial multiethnic, polyglot states of Asia and Africa. Yugoslavia, the most extreme example in the Balkans, betrays a classic Third World syndrome in terms of interethnic antagonism and intrastate insecurity. One suspects that here, again, the time factor, coupled with the way these states were brought into being as a result of decisions largely taken by major external powers, provides the most fruitful explanation for the predicament faced by the Balkan states.[16]

Barry Buzan's distinction between "strong" and "weak" states, which accords primary explanatory power to what he calls "the variable of sociopolitical cohesiveness" (Azar and Moon, 18), is related in important ways to this difference in the time available to different categories of states to complete the twin processes of state making and nation building. This comes through clearly in his conclusion: "Building stronger states is virtually the only way in which the vicious circle of unstable states and an unstable security environment can be broken" (Azar and Moon, 40). Similarly, the emphasis placed by Azar and Moon (chap. 4) on the "software" side of national security in the Third World is an acknowledgment on their part that not enough time has been available to state makers in these countries to develop the intangible ingredients of security, including the identification of the people with the state (legitimacy) and of people with each other (integration). It is also an acknowledgment of the fact that in the absence of these intangibles, the state elites in the

[16] For an insightful analysis of state making and nation building in the Balkans, see Najdan Pasic, "Varieties of Nation-Building in the Balkans and among the Southern Slavs," in S. N. Eisenstadt and Stein Rokkan, eds., *Building States and Nations* (Beverly Hills, Calif.: Sage, 1973), 2:117–41. Pasic refers to the Balkan experience as the "missing link" and "a transition between the way nations were formed in Europe at the beginning of the present era and the nation-building now going on in the developing countries" (p. 118). Pasic also makes another interesting comparison between the way the Balkan states were carved out and spheres of influence established among them and the way many Third World states were formed as a result of intraimperial understandings:

> From the Holy Alliance and the Congress of Berlin to the Yalta Conference, where spheres of influence in the Balkans were calculated in percentages, the Balkan peoples had their destinies carved out by others. The parceling out of political and national structures in the Balkans was in a substantial part the product of such external forces. In this respect, the historical circumstances surrounding nation-building in the Balkans bear a close resemblance to those in which nations and independent national states have taken shape in other parts of the economically underdeveloped world. (p. 130)

Third World are bound to take frequent recourse to the "hardware' instruments of security, namely, military force, to meet what are essen tially political challenges from disaffected groups within their popula tions.

It should be noted here that the Latin American case in terms of th availability of time for purposes of state building appears to be somewha different from that of the rest of the Third World because the forme colonies of Spain and Portugal in South America acquired political in dependence over a hundred years before the process of decolonizatior began in earnest in Asia. For a number of reasons, however, their pro cesses of state making and nation building remained retarded. Promi nent among these was the importation, along with Spanish and Portu guese colonists, of the economic and political culture of preindustria Iberia, which led to the fossilization of Latin American political devel opment. As Skidmore and Smith have pointed out, "However much Latin America struggled, it was to remain an extension ... of the Europe that had sailed west in the fifteenth century."[17]

The era of industrial society (in terms of its demonstration effec rather than its realization) caught up with Latin America about the same time that it did with much of Asia, if not Africa. (It needs to be pointed out here that the model of the industrial society includes as its essential elements a socially mobile population, forming part of a society that is culturally relatively homogeneous, that is encompassed within a legitimate state structure with adequate "infrastructural power," and that is presided over by a representative government.) While the intervening century may have provided Latin America the time to consolidate state boundaries, its social and political structures retarded other aspects of state making, above all those of societal penetration and the achievement of political legitimacy both for state institutions and for ruling elites. Thus, the acquisition of formal political independence relatively early in the game gave Latin American states only marginal advantages over their Asian counterparts. In any case, a head start of a little over a century, especially in the absence of other elements conducive to state and nation building, was not much in relation to the length of time it took Western European states to complete their process of state making.[18] It

[17] Thomas E. Skidmore and Peter H. Smith, *Modern Latin America*, 2d ed. (New York: Oxford University Press, 1989), 16.

[18] For interesting analyses of the Latin American case, see Robert E. Scott, "Nation-Building in Latin America," in Karl W. Deutsch and William J. Foltz, eds., *Nation-Building* (New York: Atherton Press, 1966), 73–83; and Howard J. Wiarda, "Social Change, Political Development and the Latin American Tradition," in Wiarda, ed., *Politics and Social Change in Latin America*, 2d rev. ed. (Amherst: University of Massachusetts Press, 1982), 3–25.

is no wonder, therefore, that state elites in Latin America continue to put as much emphasis on the "hardware" instruments of internal security as do their counterparts in other parts of the Third World.[19]

Overall, the Azar-Moon analysis cooroborates the thesis that the study of the European experience of state making is very relevant to the explanation of the current security predicament of states in the Third World. This European experience, in Tilly's words,

> cost tremendously in death, suffering, loss of rights, and unwilling surrender of land, goods, or labor.... The fundamental reason for the high cost of European state-building was its beginning in the midst of a decentralized, largely peasant social structure. Building differentiated, autonomous, centralized organizations with effective control of territories entailed eliminating or subordinating thousands of semiautonomous authorities.... Most of the European population resisted each phase of the creation of strong states.[20]

The applicability of this description to the present reality within most Third World states is too uncanny to be purely coincidental.

While this similarity between the early European and current Third World experiences of state making provides part of the explanation for the internal security problems faced by Third World states, the difference in the pace of state making and nation building and the telescoping of these two processes into a combined and drastically shortened process in the case of the Third World provides the rest of the explanation. This is the result of the fact that unlike the centuries available to most European (especially West European) state makers to complete their process of state making, today's Third World state makers are under tremendous pressure to complete this extremely complicated and costly process in only three or four decades rather than three or four centuries. As a result, the process of "primitive central state power accumulation"[21] has to be speeded up tremendously. The various phases of state and nation

[19] Alain Rouquie comes close to tackling this question while attempting to provide explanations for the military's involvement in Latin American politics, but he inexplicably shies away from addressing it directly in his otherwise knowledgeable treatise *The Military and the State in Latin America*, trans. Paul E. Sigmund (Berkeley: University of California Press, 1987).

[20] Tilly (fn. 15), 71. This conclusion is also borne out by the historical evidence presented by Youssef Cohen, Brian R. Brown, and A. F. K. Organski in their article "The Paradoxical Nature of State Making: The Violent Creation of Order," *American Political Science Review* 75, no. 4 (1981), 901–10. They argue that "instead of indicating political decay, violence in these [new] states is an integral part of the process of the accumulation of power by the national state apparatus" (p. 909).

[21] "Many of the new states of today are engaged in struggles whose logic is similar to that of the European period of primitive central state power accumulation"; Cohen, Brown, and Organski (fn. 20), 902.

building, which were undertaken and completed by and large sequentially (although with significant degrees of overlap between the phases)[22] and without any significant amount of premeditation in the case of early modern Europe, have to be undertaken and completed deliberately and simultaneously within a time-bound framework of ridiculously short duration.

This drastic shortening of the time frame and the telescoping of the various phases of state making, combined with the initially low level of state power from which state making takes place,[23] provide the primary explanation for the sharp internal challenges to the centralizing state structures in the developing countries and for the high level of violence endemic in the current phase of state making in the Third World. These challenges—whether posed in the garb of ethnicity or class or a combination of the two—and the violent responses to them are functions of the low level of legitimacy enjoyed by most Third World states within their societies; they form the core of the security problems facing these states and their regimes. Several contributions to the Azar-Moon volume as well as the second chapter of Thomas's book refer to the connection between low level of legitimacy and internal security problems in Third World states. But they do not adequately probe these linkages and do not delve into root causes that are embedded in the process of state making in the Third World. This is an area that can prove to be very fruitful in terms of further research on the interconnections between the factor of time, the process of state making, and the problem of insecurity in the countries and regions of the Third World.

IV

The security problems of Third World states are exacerbated by the fact that state making in the Third World does not take place in an interna-

[22] Stein Rokkan, in a very incisive essay in which he attempted to construct a paradigm explaining the various dimensions of state formation and nation building in Europe, provided four sequential phases over which these twin processes took place and termed them penetration, standardization, participation, and redistribution. For details, see Rokkan, "Dimensions of State Formation and Nation-Building: A Possible Paradigm for Research on Variations within Europe," in Tilly (fn. 15), esp. 572–74. Rokkan analyzed the internal variations in the patterns of nation-state building in Europe and concluded that, despite these differences within the European experience, "what is important is that the Western nation-states were given a chance to solve some of the worst problems of state-building before they had to face the ordeal of mass politics" (p. 598).

[23] This point is made by Cohen, Brown, and Organski (fn. 20), who argue that "the extent to which an expansion of state power will generate collective violence depends on the level of state power prior to that expansion ... the lower the initial *level* of state power, the stronger the relationship between the *rate* of state expansion and collective violence" (p. 905).

tional vacuum. While the internal or intrastate dimension of state making may be the primary preoccupation of state elites in the Third World, the impact of international forces, whether military, political, economic, or technological, makes a substantial and substantive difference to the fortunes of the state-making enterprise and to the larger security problematic of Third World states. This is particularly so in the contemporary era when the technologies of communication and destruction link the various parts of the world in a way that is qualitatively different from the situation prevailing in any previous historical epoch.

Moreover, as a result of the colonial experience of most Third World societies, external factors have traditionally had a predominant influence in shaping their polities and, therefore, their security environments. In fact, it would not be wrong to say that many Third World states, particularly in Africa and the Middle East but also elsewhere in Asia, emerged into the postcolonial era as sovereign entities with recognized boundaries only because they had been consolidated into separate colonial protostates by the European imperial powers in the nineteenth century.[24]

This has had two major consequences for both the internal and the external security of Third World states. First, decisions taken by colonial powers for reasons of administrative convenience or intraimperial tradeoff have been largely responsible for the ethnic mix inherited by many postcolonial states as well as for the creation of new communal identities in some instances. The colonial inheritance thus fundamentally determined the internal cohesiveness of most Third World states during their initial and crucial stages of state building and, therefore, the intensity of internal challenges to their boundaries and institutions.[25] Second, decisions taken by colonial powers have also been responsible for creating

[24] There is a growing literature on this subject, particularly in relation to Africa. A recent, perceptive article on the creation of colonial protostates in Africa is Jeffrey Herbst, "The Creation and Maintenance of National Boundaries in Africa," *International Organization* 43 (Autumn 1989), 673–92. For the creation of protostates in the guise of mandates in the Middle East, see Elizabeth Monroe, *Britain's Moment in the Middle East, 1914–1956* (Baltimore, Md.: Johns Hopkins University Press, 1963); and for the impact of the European division of Arab lands of the Ottoman Empire on international and regional security, see David Fromkin, *A Peace to End All Peace: Creating the Modern Middle East* (New York: Henry Holt, 1989).

[25] For an insightful analysis of the colonial inheritance and its impact on Third World "stateness," see Joel S. Migdal, *Strong Societies and Weak States: State-Society Relations and State Capabilities in the Third World* (Princeton: Princeton University Press, 1988). See also Crawford Young, "The African Colonial State and Its Political Legacy," in Donald Rothchild and Naomi Chazan, *The Precarious Balance: State and Society in Africa* (Boulder, Colo.: Westview Press, 1988); and Sheldon Gellar, "State-Building and Nation-Building in West Africa," in Eisenstadt and Rokkan (fn. 16), 2:384–426. For examples of the creation of "traditional" authority structures as well as the evolution of new communal identities during colonial rule, see Migdal (pp. 97–141); and Ulf Himmelstrand, " 'Tribalism,' Regionalism, Nationalism, and Secession in Nigeria," in Eisenstadt and Rokkan (pp. 427–67).

many postcolonial interstate conflicts: (1) by dividing ethnic groups into more than one state and thereby igniting the embers of irredentism, as in the Horn of Africa; (2) by denying self-determination to certain ethnic groups like the Kurds, who possibly qualified for statehood better than many that were granted that status; and (3) by leaving behind extremely messy situations, as in Palestine and in Kashmir, that have contributed tremendously to regional tensions and conflicts in the Middle East and South Asia, respectively, during the last four decades.

Equally important, in terms of the feelings of insecurity that are very widespread among Third World state elites, is the legacy of the colonial entities' individual and collective weakness and vulnerability in relation to the metropolitan centers. This sense of insecurity has been transferred after decolonization to the sphere of the Third World's relationship with the industrialized states in general and with the superpowers in particular and has largely become a function of the glaring disparities in economic, technological, and military power between the developed states on the one hand and the Third World on the other.[26] It has been further exacerbated by the division of the globe into a relatively secure and conflict-free zone, populated by European and North American states plus Japan, and the Third World, where conflict is endemic. As Ball has pointed out, "All interstate wars since the end of World War II have taken place in the Third World, although there have been industrialized country participants in some of these conflicts" (p. 33).[27] In fact, some analysts have argued that conflict in the Third World has until recently been encouraged by superpower policies largely aimed at testing each other's political will and power projection capabilities in those areas of the globe that are not of vital concern to either superpower and, therefore, do not threaten the maintenance of the central strategic balance.[28]

This de facto division of the globe, roughly corresponding to the core-

[26] It is worth pointing out in this context that Krasner's (fn. 3) assumption that "political weakness and vulnerability are fundamental sources of Third World behavior" (p. 3) is substantially correct. However, his characterization of the consequent North-South relationship as "structural conflict" is, like its obverse, dependency theory, too extreme, simplistic, and one-dimensional in nature; it does not do justice to the much more complex reality of that relationship.

[27] The high incidence of violent conflict in the Third World is borne out by a number of studies, including Mark Zacker, *International Conflicts and Collective Security* (New York: Praeger, 1979); and Nazli Choucri, *Population and Conflict: New Dimensions of Population Dynamics*, Policy Development Studies No. 8 (United Nations Fund for Population Activities, 1983). For an earlier, pioneering study of the subject, see Istvan Kende, "Twenty-five Years of Local Wars," *Journal of Peace Research* 8, no. 1 (1971), 5–22.

[28] This point was best made by Sisir Gupta two decades ago; Gupta, "Great Power Relations and the Third World," in Carsten Holbraad, ed., *Super Powers and World Order* (Canberra: Australian National University Press, 1971), 105–39.

periphery dichotomy of the world system theorists,[29] allows for the exportation of the developed world's conflicts to the Third World, while effectively insulating the "core" of the international system from the conflicts and instabilities prevalent in the Third World. As a result, it enhances the insecurity of Third World state elites who suffer from a feeling of dual impotence. First, they are unable to prevent superpower rivalries and conflicts from penetrating their polities and regions, and second, they are equally unable to affect, except marginally and in selected cases, the global political and military equation between the two superpowers and their respective alliances. This conclusion is borne out by even a casual reading of the Litwak-Wells volume and is corroborated by Thomas's statement that "the outlook for the Third World remains bleak. While it is very far from true to suggest that everything that happens to them is a result of external factors, it is fallacious to believe that indigeneous factors play the most influential role most of the time" (p. 199).

All this does not mean, however, that external threats of conflict and *intervention in relation to Third World states do not also arise from* within their regions, that is, from other Third World states. Obviously, Third World regions do possess autonomous dynamics of conflict and cooperation. In fact, the predominant reality of these regional dynamics is the great propensity for conflict that inheres within them. As the various contributions to the Litwak-Wells volume demonstrate, there are reasons intrinsic to Third World regions, for example, historical mistrust, territorial disputes, ethnic overlap, and hegemonic ambitions, that provide much of the raw material for interstate conflict in the Third World. Two points are worth noting in this context, however. First, many of these intrinsic reasons for intra–Third World conflict are related to, if not the products of, external domination during the colonial era. Second, the permissive attitude on the part of the dominant global powers toward conflict in the Third World promotes and exacerbates interstate violence in the gray areas of the globe.

V

One major factor that has increased the level of violence in the developing world as well as the propensity of Third World states to indulge in

[29] For details of the core-periphery dichotomy, see Johan Galtung, "A Structural Theory of Imperialism," *Journal of Peace Research* 8, no. 2 (1971), 81–117; and the various works of Immanuel Wallerstein.

interstate conflict is the transfer of modern weapons and weapons technology from the industrialized countries to various parts of the Third World. According to one specialist on the arms trade, "These transfers have resulted in a significant shift in military resources from the industrialised 'North' to the underdeveloped 'South,' producing new configurations of power and contributing to the intensity and duration of regional conflicts."[30] Certainly, weapons are mainly of instrumental value and are not in themselves the primary cause of war; but relatively sophisticated weapon systems that provide a Third World state with temporary technological superiority over a regional rival are very often a crucial factor in the calculation of decision makers to escalate disputes to a point where war becomes a distinct possibility. To cite just one instance: Pakistan's decision to go to war against India in 1965 with the objective of changing the status quo in Kashmir was based in part on the assessment that the former's edge in sophisticated weapons over the latter was likely to be eroded over the next few years.[31]

Sophisticated weapons acquired at great cost to provide greater security can often increase prospects of conflict and, therefore, add to the insecurity of Third World states.[32] Andrew Ross's contribution to the Azar-Moon volume (chap. 7) concentrates on "the various forms of arms acquisition options available to Third World countries and the impact of alternative acquisition strategies upon national security" (p. 154) and makes the point sharply that "the acquisition of military power may itself erode rather than enhance security" (p. 153). As a result of the combination of various factors (including the escalation in military technology, the superpowers' strategy to use arms transfers as political instruments to buy the loyalties of Third World clients, the leading arms exporters' interest in using arms sales as a major booster for their economies, and the inability of even the most technologically advanced Third World states independently to manufacture more than a fraction of the sophisticated weapons they need or desire to possess), "the Third World's de-

[30] Michael T. Klare, "The Arms Trade: Changing Patterns in the 1980s," *Third World Quarterly* 9 (October 1987), 1257.

[31] Gowher Rizvi, "The Rivalry between India and Pakistan," in Barry Buzan et al., *South Asian Insecurity and the Great Powers* (Basingstoke: Macmillan, 1986), 107–8.

[32] The availability of large surplus stocks of modern weaponry combined with cold war motivations on the part of superpower suppliers has contributed to regional arms races and to the instability of regional balances, which must constantly be restabilized at higher levels of technological sophistication. As Raju Thomas has pointed out in the case of South Asia, "The net result was that both India and Pakistan acquired substantially more arms than they otherwise would have thus producing less regional security for both states at a much higher price"; Thomas, "Strategies of Recipient Autonomy: The Case of India," in Kwang-Il Baek, Ronald D. McLaurin, and Chung-in Moon, eds., *The Dilemma of Third World Defense Industries: Supplier Control or Recepient Autonomy?* (Boulder, Colo.: Westview Press, 1989), 188.

pendence upon arms imports from the advanced industrial countries" has become "the defining characteristic of post-colonial North-South military relations" (Ross, in Azar and Moon, 156).

The 1980s saw an appreciable increase in the capacity of certain Third World states, like India and Brazil, to produce, and even export, indigenous arms. There was also a perceptible decline in the value of arms deliveries to the Third World during most of the 1980s as a result of the fall in the price of oil, which drastically reduced the purchasing capacity of major oil-exporting countries that had been among the leading acquirers of sophisticated weaponry, and of the saturation of many Third World markets. However, these two phenomena mask a different, and growing, form of weapons transfer from developed to developing countries: the transfer of sophisticated arms production technology, now an integral part of the international arms trade.[33] The transfer of such technology has two major consequences. On the one hand, as conventional wisdom holds, given both the high rate at which weapons and weapons technologies become obsolete in the late twentieth century and the inability of Third World countries to keep up with the latest technologies, the transfer of weapons technology amounts to nothing more than the substitution of one form of dependence for another. Indeed, it might even increase the level of dependency. On the other hand, as Ross has argued in the Azar-Moon volume,

> The nature of military dependence undergoes a subtle but potentially profound transformation as developing countries turn from arms imports to arms production. . . . A static dependence relationship is inevitable when a country relies upon foreign arms suppliers. But when arms production programmes are initiated, and military production technology rather than arms are imported, a more dynamic relationship is established, one that has an inherent potential for the reduction, if not elimination, of military dependence. (pp. 169–70)[34]

The political autonomy of Third World arms recipients was further enhanced during the 1980s by the appreciable increase in the number of arms suppliers and the increasingly intense competition among them, the decline in the market share of the superpower suppliers, and the increasing transformation of the nature of the arms market from a seller's to a

[33] For a detailed discussion of this issue, see Michael T. Klare, "The Unnoticed Arms Trade: Exports of Conventional Arms-Making Technology," *International Security* 8 (Fall 1983), 68–90.

[34] Ross's contention is borne out by Raju Thomas's study (fn. 32) of Indian weapons procurement and production policy, which, according to its author, attempts "to strike an optimum balance among the three basic strategies of indigeneous production, licensed production and overseas purchases" (p. 199).

buyer's market. However, this should not lead us to conclude that commercial calculations alone dictate the flow of arms supply or of weapons technology from the industrialized countries to the Third World. There are a number of political and strategic considerations, many of them connected with superpower global rivalries and with the myriad of political and strategic links between the superpowers on the one hand and the leading nonsuperpower arms suppliers on the other, that have a major bearing on the pattern of arms trade and even more on the transfer of sophisticated weapons technology to Third World recipients.[35]

Either way, the net effect on the overall security of the Third World can turn out to be negative. If the transfer of weapons technology increases the dependence of Third World states on major industrial powers, then the feeling of vulnerability and insecurity among the elites of Third World states is intensified. However, any reduction in such dependence, consequent upon the transfer of technology and the diversification of sources of arms supply, increases the autonomy of decision making in relation to war and peace as far as the more developed Third World countries are concerned, and it removes important international constraints on their conflictual behavior. Transfer of weapons technology also strengthens their war-fighting capacity by making them relatively independent of the original suppliers for spare parts and ammunition and by increasing the sophistication of the technology that their war machines can command indigenously, at least in the short run. As a result, wars in the Third World, especially among major regional actors, can now be started without the protagonists being overly concerned about supplier reactions, can be sustained for longer periods of time, and can be far more costly in human and material terms than they were in the past decades.

Nuclear proliferation in the Third World is a subset of the problems connected with the transfer of sophisticated weapons technology. It is, however, the most dramatic among this set of problems. As the only Third World security issue that ties Third World security concerns directly to those of global security, it is the only one in which the great powers have taken direct and immediate interest. They have attempted to institutionalize international controls on Third World behavior through the medium of the Nuclear Non-Proliferation Treaty (NPT). However, as Thomas has pointed out, "From the point of view of Third World states (even those that have joined the NPT), the nuclear non-proliferation regime in its present form institutionalises inequality between

[35] For details of the latter argument, see Stephanie G. Neuman, "Arms, Aid, and the Superpowers," *Foreign Affairs* 66 (Summer 1988), 1044–66.

nuclear-weapon states and non-nuclear-weapon states" (p. 141). This has led to a certain amount of tension between the members of the nuclear club, especially the superpowers, and the have-nots in the nuclear arena. Such leading members of the latter group as India, Pakistan, Israel, and South Africa have, in fact, attempted to circumvent the controls imposed by the nuclear club in order to expand their actual or potential nuclear capabilities. These capabilities, all shrouded in policies of deliberate ambiguity, include a substantial arsenal of small nuclear weapons (Israel), a "peaceful" underground nuclear explosion coupled with an increasingly sophisticated delivery capability (India), a dual-track effort to manufacture nuclear warheads by uranium enrichment and/or plutonium reprocessing (Pakistan), and an atmospheric explosion that seemed suspiciously akin to an atomic test (South Africa).[16]

Such unacknowledged but nonetheless credible instances of nuclear proliferation pose problems not merely for the security of Third World states and regions but for the security of the international system as a whole. Moreover, the problems are not confined to the largely abstract ones of managing a world with a dozen or so nuclear powers. The practical problems of proliferation are far more acute because the four de facto nuclear powers mentioned above are all involved in regional conflicts and confrontations that could become overtly nuclear[17] and consequently lead to the direct involvement of one or both of the superpowers in the disputes.

It is, however, impossible to put the genie of nuclear proliferation back into the bottle, especially because possession of nuclear weapons has become the hallmark of enhanced status within the international system. The Chinese example has very sharply driven home this lesson to Third World ruling elites, in particular those of the larger and regionally powerful states. In light of the Chinese experience, no Third World leadership aspiring to graduate to the status of a major, or even moderately influential, actor in the international system can feel comfortable about giving up its nuclear option. This factor of prestige, combined with genuine security concerns on the part of several Third World states facing

[16] For overviews of the four countries' nuclear capabilities, see, for Israel, Peter Pry, *Israel's Nuclear Arsenal* (Boulder, Colo.: Westview Press, 1984); for India and Pakistan, Carnegie Task Force on Non-Proliferation and South Asian Security, *Nuclear Weapons and South Asian Security* (Washington, D.C.: Carnegie Endowment for International Peace, 1988); and for South Africa, Michele A. Flournoy and Kurt M. Campbell, "South Africa's Bomb: A Military Option?" *Orbis* 32 (Summer 1988), 385–401.

[17] For a prospective scenario on the Indian subcontinent, see Leonard S. Spector, "India-Pakistan War: It Could Be Nuclear," *New York Times*, June 7, 1990, p. A23. For the Middle East, where a situation of de facto nuclear monopoly prevails, see Helena Cobban, "Israel's Nuclear Game: The U.S. Stake," *World Policy Journal* 5 (Summer 1988), 415–33.

potential anatagonists that are nuclear or near-nuclear powers, has created a situation in which the security of a number of leading Third World states has become intertwined with the issue of nuclear proliferation. While all major Third World states on the nuclear threshold continue to abide by a policy of deliberate ambiguity for the moment, external or domestic stimuli could change this situation with dramatic suddenness in individual countries, thereby setting off chain reactions that may be difficult to control.

VI

As security cannot be bought cheaply in the late twentieth century, many Third World states spend substantial proportions of their relatively meager resources on the security sector.[38] While expenditure on costly weapon systems and military technologies is a part of security expenditure, Ball has demonstrated that it is not the major part in the case of the overwhelming number of Third World states, which "appear to spend a very high proportion of their security budgets on operating costs, particularly salaries and emoluments for the troops" (p. 393). Appendix 1 of Ball's book (pp. 396–402) examines the evolution of operating costs as a percentage of total security expenditure of twenty selected Third World countries for the years 1951–79 and finds that in all cases except one (Iran under the Shah) operating costs have clearly dominated in the mix of security expenditures for these countries. Ball concludes therefore that while "in the public mind, security expenditure in the Third World is firmly linked with the arms trade" (p. 107), "for most of the Third World, the arms trade and security expenditure are not synonymous: Operating costs, particularly personnel-related outlays, form a large and permanent portion of most developing countries' security budgets" (p. 111). This, Ball suggests, is connected to the fact that "the internal security role of the armed forces is considerable throughout the Third World and, in many cases, is their primary function" (p. 393).

Ball's sample leaves out such oil-rich and population-poor countries as Saudi Arabia and Kuwait, for which capital costs, including expenditure on defense infrastructure and weapons procurement, would be appreciably higher than is reflected in her data. Nevertheless, her basic point regarding high operating costs and their relationship to internal regime

[38] Ball has correctly pointed out that it is preferable in this context to refer to the "security" sector rather than the "military" sector, "in order to indicate the inclusion of paramilitary forces" and to reflect "the fact that Third World governments frequently use their armed forces to maintain themselves in power, that is, to promote regime security" (p. xvi n. 2).

and state security is valid for most Third World countries. This reflects not merely the relatively low level of technology and the high level of manpower required by Third World states for the maintenance of internal control; it is also indicative of where they are along the continuum of the state-making enterprise. As the European experience has demonstrated, three areas—taxation (extraction of resources under the protection of coercive state agencies), policing (maintaining domestic order where it has already been imposed), and warfare aimed at the primitive accumulation of state power (extending and consolidating a particular political order by the use of force against potential as well as dissident subjects and fending off rival claimants to the same territorial and demographic space)—comprise the bulk of the activities undertaken by early state makers.[39] These are all labor-intensive tasks that engage relatively large numbers of persons in the security arena and thereby raise the ratio of operating (including personnel) costs relative to the total expenditure on security. Security sector costs in the Third World are understandably linked to the performance of these essential functions in their current early stage of state making.

In this context, Ball's central question regarding the relationship between security expenditure and development seems to be of secondary concern, if not misplaced. Despite the declared commitment of Third World state elites to the goal of development (defined as economic growth plus some degree of distributive justice), as far as most of them are concerned this is an instrumental value that helps them achieve their primary objectives of political legitimacy and state and regime security. Therefore, Ball's conclusion that "available evidence does suggest that expenditure in the security sector is more likely to hinder than to promote economic growth and development in the Third World" (p. 388) misses the essential motivation behind such expenditure, even though the point may be valid.

This motivation has to do primarily with the "primitive central state power accumulation" mentioned above and secondarily with meeting threats from the regional environment. As a result, development, measured as a serious objective and not merely on the basis of the rhetoric of Third World leaders, comes a poor third in the policy priorities of most Third World elites and is hardly ever considered an autonomous goal that deserves to be fulfilled independently of security considerations. It

[39] For details, see the following essays in Tilly (fn. 15): Samuel E. Finer, "State- and Nation-Building in Europe: The Role of the Military" (chap. 2); Rudolf Braun, "Taxation, Sociopolitical Structure, and State-Building: Great Britain and Brandenburg-Prussia" (chap. 4); and David H. Bayley, "The Police and Political Development in Europe" (chap. 5).

is not surprising, therefore, that, as Ethan Kapstein has argued in chapter 6 of the Azar-Moon volume, "Third World states have allocated scarce resources to meet national security threats, and in so doing have influenced the timing and/or trajectory of economic development" (p. 138).

Furthermore, given the fragility of many Third World polities, it is no wonder that "one negative effect of security expenditure," as Ball terms it, has been "the strengthening of the armed forces at the expense of civilian groups within society" (p. 390). There is no other institution that is more important as far as the interface between issues of state making and those of internal and external security are concerned than the military. While Ball devotes the whole of chapter 1 to the discussion of the military's role in development and offers a critical evaluation of various theories that profess to explain this role, none of the four volumes under review has attempted to relate the role of the military to the process of state and nation building in the Third World. The closest one comes to it is in the last few pages of Ball's volume:

> By relying on the armed forces to remain in power or by producing political and economic conditions that provide the military with the justification for intervention, many governments have facilitated the entry of the armed forces into the political arena." (p. 391)

This generalization glosses over the fact that Third World polities are currently caught in a vicious circle that is a product of their historical circumstances. As the early modern European experience has demonstrated, the role of the coercive apparatuses of state—meaning primarily military and paramilitary institutions—in the early phase of state making is considerable. In the case of most Third World states the problem has been compounded by the existence and combination of two additional factors. The first is the weakness of civil society and of other political institutions, which precludes the emergence of strong checks on the natural proclivity of the security apparatuses to usurp as much of the power and resources at the command of the state as possible. Second, the encapsulation of the various phases of state and nation building into one all-encompassing phase and the drastic curtailment of the time available to Third World states for the completion of these twin processes enhance the political importance of the coercive functions of the state and, therefore, of the agencies that perform such functions. Even in India, where a democratic political system has operated more successfully than elsewhere in the Third World, the important and increasingly dominant role of the security apparatuses is clearly visible in states like Punjab and Kashmir, which pose major overt challenges to the Indian state in the

arena of state and nation building. It should therefore come as no surprise that the security sector in most Third World states hogs a large share of the state's disposable resources irrespective of the impact this may have on the process of economic development.

Furthermore, in terms of the allocation of scarce resources to the security sector, there seems to be very little difference between those Third World polities that are overtly dominated by the military and those that are under civilian control. This is demonstrated by Ball's own data (Figure 10-1, p. 387), which include several states under civilian rule within the category of the heaviest spenders on security. It is also corroborated by a recent study of defense spending by the member states of the Association of Southeast Asian Nations (ASEAN), which, notwithstanding several caveats and qualifications, came to the conclusion that "the countries [in ASEAN] in which the military has the largest political role (Thailand and Indonesia) are the ones in which defence spending has grown more slowly than the ASEAN average."[40] State making and the violence that accompanies it obviously make no distinction between military-dominated and civilian-ruled polities in the Third World.

VII

Despite their divergent treatments of the subject and some of their shortfalls noted above, the four volumes reviewed in this article taken together make a substantial contribution to the study of Third World security problems by highlighting areas of analysis that have remained relatively neglected so far. These problems will remain with us for the foreseeable future in spite of the changes that seem to be underway in the superpowers' relations with each other and with the rest of the world. In fact, recent and projected changes may even contribute to the accentuation of some of the security problems faced by Third World countries and regions. The anticipated withdrawal of one superpower from the Third World arena may not turn out to be an unmixed blessing. The other superpower may feel free to act more cavalierly as far as the security and the vital interests of Third World states are concerned; it may be tempted to intervene militarily if developments in what it considers to be "strategic regions" of the Third World are not to its liking. Important Third World state elites, deprived of the presence of a balancing

[40] David B. H. Denoon, "Defence Spending in ASEAN: An Overview," in Chin Kin Wah, ed., *Defence Spending in Southeast Asia* (Singapore: Institute of Southeast Asian Studies, 1987), 49.

power that could in some measure neutralize the dominant superpower's interventionist proclivities, may therefore begin to feel more vulnerable and insecure. If this happens, such an escalation of insecurity will be reflected in their internal and external behavior patterns.

It could also happen that a genuine disentanglement on the part of both superpowers from arenas of tension and conflict in the Third World may remove some of the restraints on the conflictual behavior of important Third World states. The aggressive potential of those states has been constrained by the apprehension that it could draw negative reactions from one or both of the superpowers and thereby end up tipping the regional balance against them. But were the superpowers to pull back, it might lead to greater assertiveness on the part of regionally preeminent powers interested in translating their preeminence into hegemony or at least into a managerial role within their respective regions. Resistance by other countries in a particular region to such hegemonic behavior might in turn, lead to situations of violent interstate conflict relatively unhindered by concerns regarding superpower intervention.

Furthermore, the prospects of conventional arms control pacts and troop reduction agreements between the superpowers, which are expected to lead to major redundancies in their arsenals, are already spurring both Washington and Moscow to increase arms sales abroad.[41] This trend can be expected to accelerate once these pacts become realities and force both superpowers to remove various categories of conventional weaponry, including tanks, artillery, aircraft, and helicopters, from their inventories. Much of this surplus hardware is expected to be sold to Third World countries to fulfill hard currency needs (a particularly important consideration for Moscow), to shore up friends and allies by making them more "self-reliant" in terms of hardware, and to find alternative sources of profitable returns for domestic arms industries. This projected escalation in the transfer of sophisticated weaponry to the Third World in the 1990s will almost certainly reverse the trend of decreasing arms transfers to the developing countries that had been visible for most of the 1980s[42] and will further enhance the destructive potential of Third World conflicts.

In the final analysis, however, most of the deep-seated sources of conflict and violence in the Third World—sources that inhere within Third

[41] For details, see Robert Pear, "Prospects of Arms Pacts Spurring Arms Sales," *New York Times*, March 25, 1990, p. 12.

[42] For the latest analysis of this trend, see Richard F. Grimmett, *Trends in Conventional Arms Transfers to the Third World by Major Suppliers, 1982–1989* (Washington, D.C.: Congressional Research Service, Library of Congress, 1990), esp. 1–3.

World societies and are related to their simultaneously ongoing processes of state making and nation building—cannot and will not be fundamentally determined by superpower actions and interactions, even if the latter has had the capacity to exacerbate many of them in the postwar era.[43] Therefore, although changes in superpower relations may continue to affect some of these sources of conflict and insecurity in the Third World, these changes alone are not capable of transforming the basic nature of the security predicament of the Third World states. As it stands, the existing parameters of the security problematic of the Third World can be altered only if Third World states have adequate time to complete the twin tasks of the state making and nation building, plus enough political sagacity on their leaderships' part to attempt to accomplish these tasks in as humane a manner as possible. At such time the security concerns of developing states will approximate more closely those of the developed states, which in the traditional literature on international relations have constituted the model for state behavior in the security arena.

[43] Notwithstanding the fact that the most highly visible facet of the latest Gulf crisis has been its global dimension involving the United States projection of power in the Gulf, this crisis has its origins in the internal dynamics of the region. These, in turn, are closely intertwined with issues regarding the establishment and legitimization of state boundaries, institutions, and regimes in the Middle East in general and the Persian Gulf in particular. Iraqi ambitions regarding Kuwait date back to the founding of the Iraqi state under British tutelage in the aftermath of the disintegration of the Ottoman Empire following World War I. Iraq's claims on Kuwait rest both upon Ottoman assertions of sovereignty over Kuwait and upon the widespread feeling in the Arab world that post-Ottoman borders in the Fertile Crescent and the Gulf were arbitrarily drawn by Western colonial powers to suit their own selfish requirements and are therefore less than fully legitimate. In recent times Baghdad made two abortive attempts, in 1961 and 1973, to enforce its territorial claims on Kuwait. Viewed in its proper historical perspective, the latest crisis is therefore not exclusively an Iraqi attempt to control Kuwait's huge oil resources and dictate oil production and pricing policies within OPEC. While the American reaction to the Iraqi invasion of Kuwait in August 1990 globalized the crisis, it was basically just that—a *reaction* to a crisis that was fundamentally grounded in regional realities and intimately related to rival claims over both territorial and demographic space (and, of course, over the only major resource of the region), as well as linked to issues of state and regime legitimacy in the Arab littoral of the Persian Gulf.

REFORM AND THE REDEFINITION OF THE SOCIAL CONTRACT UNDER GORBACHEV

By JANINE LUDLAM

Anders Aslund, *Gorbachev's Struggle for Economic Reform*. Ithaca, N.Y.: Cornell University Press, 1989, 219 pp.
Walter D. Connor, *Socialism's Dilemmas: State and Society in the Soviet Bloc*. New York: Columbia University Press, 1988, 299 pp.
Jerry Hough, *Russia and the West: Gorbachev and the Politics of Reform*. New York: Simon and Schuster, 1988, 301 pp.
Moshe Lewin, *The Gorbachev Phenomenon: A Historical Interpretation*. Berkeley: University of California Press, 1988, 176 pp.

BY becoming president of the Soviet Union in March 1990, Mikhail Gorbachev skillfully managed to decrease the power of the ruling bodies of the Communist Party and to transfer it to the newly created executive structure of the state government. The power of the party's top echelon was further reduced in July 1990 when a new Politburo was elected at the 28th Party Congress. The Politburo was swept clean of its old membership with the exception of Gorbachev himself (who was easily elected general secretary by the Congress delegates) and V. A. Ivashko (now a deputy to the general secretary), and its size was expanded, in part to include a representative from every Soviet republic. Gorbachev thus freed himself of many policy restraints placed upon him by the Politburo, while the party—the majority of whose members, despite factionalism and defection, are still loyal to Gorbachev—remains the dominant political force in Soviet society.

As president, Gorbachev faces no serious opposition in the legislative branch of the government for the time being. Although the infamous Article 6, which guaranteed the Communist Party's monopoly on power, was struck from the Constitution last March, no opposition party has yet gathered enough strength to challenge the communists' real monopoly on power, which extends from the villages to the ministries. The vast majority of the peoples' deputies in the Congress and the Supreme Soviet of the Soviet Union are members of the Communist Party: of the 2,250 deputies, 1,973 hold party cards.[1] Thus, by shifting some power from the

[1] *Narodnye Deputaty SSSR: Spravochnik serii 'Kto est kto'* [People's deputies of the USSR: Reference book of the "Who's Who" series] (Moscow: Vneshtorgizdat, 1990).

party elite to the government, Gorbachev has amassed a large amount of power in his own hands.

Although Gorbachev appears to be unafraid to use traditional methods of power politics to manipulate his country's political structures, he has employed a more tactful approach toward the population at large. His understanding of the social dynamics and the structure of Soviet society is more profound than that of his predecessors. Appreciative of the changes that have taken place in the makeup of the Soviet population over the last forty years, Gorbachev seeks to accommodate these changes within a partially revised political and social structure in the hope that the result will be a more efficient and productive society. How successful is Gorbachev likely to be? While reviewing some recent books on this important theme, I shall analyze the forces that are propelling his efforts at economic and political reform as well as those that may obstruct them.

Background

In the years following Stalin's death, Soviet leaders sought to redefine the legitimacy of the Soviet polity and the control of Soviet society. They aimed to get away from terror and arbitrary lawlessness (where legal boundaries were deliberately fluid) and to move toward a more "contractual" political legitimacy and social control, based primarily on economic guarantees to the population. Still, they counted on the continued deprivation within mass society of real political power and political concerns, and on the continued dominance of an official state society over an indigenous civil society. This rather limited social contract served several purposes, for a time. It was undoubtedly a relief from the terror and arbitrariness of the Stalin period. It also fit well into the extensive phase of industrial development, in which most workers held basic blue-collar production jobs, were nonintellectual, and were more interested in economic security than in political rights and cultural expression. This contract neglected only a minority of the population: intellectuals, highly skilled technical workers (engineers, chemists, and so forth), as well as the emerging service class.

As the economy and society developed, the limitations of this Khrushchev-Brezhnev social contract grew more conspicuous, and also more dangerous to ignore. The relative ineffectiveness of the extensive approach in the emerging postindustrial phase of economic development revealed that a more modern, intensive strategy for further development was needed: a strategy based on better-skilled and better-educated workers (manual, managing, designing, and intellectual) who had been pour-

ing out of schools and institutes at an increasing rate during the previous decades, but who had found their skills and talents underutilized. This had political implications: higher-skilled labor had become a crucial ingredient for further development, yet the social contract tacitly agreed upon over thirty years ago was, for the most part, geared toward lower-skilled, blue-collar workers. Although the latter's importance to the economy was still great, it was declining relative to their better-educated cousins and neighbors.

Another integral part of the social contract was the mutual agreement between the leadership and the populace that Stalinist terror and generalized lawlessness should be tempered. The masses, like their leaders, found great relief in this aspect of the contract. But the young, educated, upwardly mobile professionals in the Soviet Union today do not personally remember the Stalin days. Furthermore, they themselves did not agree to that contract. Just as the Khrushchev-Brezhnev social compact does not advance an intensive approach to economic development, it does not suit the broad class of skilled and educated workers that has become indispensable to the economy as it makes its transition from the industrial to the postindustrial world.

In this essay, I will develop three themes: (1) Gorbachev has realized that the present social contract cannot properly serve the comprehensive reforms he has visualized. (2) Pressure to change the social contract has built up within society itself, beyond the control of the Soviet regime; that is, the behavior and decisions of groups in society (concerning place of work, participation in communist ideology, higher education, diligence in work, and so on) are weakening the old contract. Moreover, the government itself has deliberately loosened its control over the behavior of the population with the end of terror, the weakening of economic incentives, and the more open flow of information from outside. This challenges the idea that in an authoritarian system the government is the only actor involved in defining the terms of the contract. (3) There will be "winners" and "losers" as a result of a revised social contract. As he moves forward with reform, Gorbachev will continue to face the resistance of those in entrenched positions (persons who have benefited from the post-Stalin contract), especially blue-collar workers and lower-level party functionaries.

Social Contract Theory and the Soviet Case

The application of social contract theory to the Soviet Union became popular (though not universally accepted) in the 1970s as a way of ex-

plaining the stability of a Soviet political system that no longer relied on terror and a charismatic absolute leader for its survival and even legitimacy. The general idea of contract theory is that both regime and population contribute something that the other needs for its survival. In a general case, the government agrees to protect its citizens and to provide circumstances for their well-being, while the population agrees to acknowledge the legitimacy of the government, at least passively. Some scholars have argued that because Soviet citizens have not participated openly in government in the past—that is, they have not been voters in the Western sense—no contract could have been reached between the Soviet regime and its populace. I maintain, however, as do others, that the Soviet regime has been forced to consider the interests of the population in order to protect its own ability to govern. In the early 1960s under Khrushchev, for example, riots over announced increases in food prices broke out in the southern part of the Soviet Union; the increases were never enacted. Brezhnev's consistent policy decisions protecting the welfare state in return for labor's quiescence and its acceptance of the Soviet regime's right to rule are evidence of an implicit agreement, or contract. For five years, Gorbachev had avoided the question of raising prices—especially on basic commodities—despite economists' insistence that such increases were absolutely necessary for the welfare of the economy. Now that the issue can no longer be ignored, the government has promised that any increase in prices will be accompanied by corresponding increases in wages, and that pensioners and the poor will receive adequate social benefits. Because fear of unrest due to a decline in living standards is very great, the government has endured massive subsidies on food and basic items so as not to disrupt economic, and potentially political, stability. I therefore make the assumption that the application of social contract theory to the Soviet case can produce insights into the relationship between state and society in the Soviet Union.

In most countries, political leaders and citizens understand that an implicit social contract—a balance (or sometimes an imbalance) between economic, political, and civil guarantees and restrictions—partially legitimizes the authority of the leadership to rule, and guarantees controlled social behavior on the part of a majority of its citizens. Thomas Hobbes, one of the first political theorists to elaborate on the concept, argued that the original social contract developed so that people could leave what he called the "state of nature"; that is, man in his pregoverned state, which was unpredictable and violent, "a war of every man against every man." In this primitive contractual stage, man renounced his natural liberty and authorized a single ruler (or oligarchy of rulers) to govern him. In

return, the ruler agreed to protect him from an untimely death (an extremely likely fate as long as man lived in the state of nature). This rudimentary contract was the source of the ruler's legitimacy. At that point man, now a "citizen," received little in return for the absolute obedience he gave the ruler. Hobbes's ruler was not accountable to the individuals who had sanctioned his power, nor did these individuals have the right to participate in the shaping of the terms of the contract. Contract theory subsequently went beyond being a theory of the historical *origin* of government/society to become a theory that, in addition, explained the *nature* of government/society.

Hobbes's description of the nature of government/society is partially obsolete now, in that the citizen has in many countries gained more rights and guarantees vis-à-vis his government. John Locke went beyond Hobbes in arguing that the political authority of a government was restricted by a citizen's inalienable rights, especially his right to have his property protected. A century later, Jean-Jacques Rousseau stated that active and continued participation by citizens in their government's political processes was an indispensable condition for the contractual relationship between a government and its people, or "government by consent."

In the advanced industrial democracies, the duties of citizens to the state have remained basically the same (compliance/support for the existing government, controlled social behavior), while the rights and guarantees given to citizens by the state have proliferated. In the United States, the right to participate in the political process has been secured, respect for the individual has been a long-standing premise of legitimacy, and the right to be better-off economically than one's neighbor (which has always elicited more support in the U.S. than has the socialist ideal of economic egalitarianism) is taken for granted. At the same time, the government has provided a limited safety net for the country's elderly and sick. Sweden has struck a somewhat neater, and potentially less discordant, balance between political, civil, and economic rights and responsibilities. While similar to the United States in the first two aspects, Sweden is attempting to guarantee a constraint on economic inequality: it allows its highly centralized trade union organization (instead of the state) to distribute investment and profits and to limit wage inequality (for the same work) throughout the country's so-called sheltered economy (that is, enterprises not involved in foreign competition).

The Soviet Union, on the other hand, has historically defined the social contract between its government and citizens in much more restricted terms. While it has emphasized the responsibility of citizens to

the government, it has given short shrift to the rights of citizens vis-à-vis the government, especially civil and political rights. In fact, one could argue that a social contract did not emerge in the Soviet Union until after Stalin's death in 1953. One might ask whether citizens were indeed protected from violence and death during the Stalin period, or whether that period was a perpetuation of the uncertain and arbitrary state of nature described by Hobbes. Was the comprehensive system of neighbors telling on neighbors in order to preempt a strike against their own persons not a "war of every man against every man," the very feature that defined the state of nature? Thus the Soviet Union was in the mid-1950s just emerging from a deliberately reimposed state of nature. Its first social contract, despite the already highly industrialized character of the Soviet state, was Hobbesian in content; it reflected the continuing dominance of state over civil society and the low (repressed) level of the political culture of the masses.

Now, in the 1990s, Gorbachev is trying to accomplish more than either Khrushchev or Brezhnev aspired to. The low level of the political culture of the masses and their escape into private life has become a hindrance. The official dominance of state over civil society has meant that a second, somewhat autonomous civil society has developed which the state cannot easily tap into and use to promote its own objectives. The "*extreme* autonomy of the state" (Walter Connor's term for the condition in which a state is not dependent on the support of social groups or classes for its existence) has left state-society relations severely underdeveloped (Connor, 9–16). Although dominated by the state, Soviet society has remained separate from it and has had to find outlets other than political channels for expressing itself. By the early 1980s, the social policy of "to each according to his needs" rather than "to each according to his labor" had resulted in what Jerry Hough calls "excessive coddling of the workers" (p. 99). It was a policy that left the workers little incentive to work hard and well, and it consciously, albeit reluctantly, sacrificed high rates of economic growth and encouraged economic inefficiency. Gorbachev's comprehensive reform package was not served well by such an underdeveloped social contract.

Pressure on the Post-Stalin Contract

Two important questions arise regarding the application of contract theory to the Soviet case. First, what concrete changes are being made in the social contract? In other words, what policy changes are being implemented, or at least visualized, that will change the content of the con-

tract? Second, is the government the only player in the shaping of tl
contract? That is, is the Soviet social contract still primarily Hobbesia
(where individuals have no voice in defining its terms), as it was und
Stalin, or is it becoming more Lockean (where individuals are permitte
some say)? It is fairly easy to find answers to these questions. The answe
to a more difficult question may shed some light on the dilemma of hov
Soviet society interacts with its authoritarian state: Who or what is re
sponsible for forcing a change in the contract? Is Gorbachev's struggl
for reform putting pressure on the old social contract, or are social force
themselves pushing for a new social contract and, implicitly, for reform?

Both Hobbes and Locke had observed that the social contract was often not freely negotiated between a government and natural men. Hobbes noted that a strongman could force individuals to participate in the contract; Locke stated that individuals could be induced into compliance by a godlike "king" (not dissimilar to Max Weber's charismatic leader). In a recent article, Peter Hauslohner comments that the word "contract" is somewhat misleading in the Soviet context, for the deliberations that lead to social policy are conducted by one side only—the government.[2] This explanation, however, does not take into account the existence of complex forces in Soviet society. Westerners in general have been too willing to see the Soviet state as all-powerful and as possessing the ability to control its citizens at will. In reality, as was noted earlier, while the state officially continued to dominate its citizens, an unofficial "second" society developed alongside—a society in which the populace privatized its existence, a second economy flourished, and samizdat culture proliferated. Soviet society, although not independent of its state, became individualized and decentralized.[3]

In *The Gorbachev Phenomenon*, Moshe Lewin argues that Western scholarship on the Soviet Union has focused primarily on the state, to the exclusion of other aspects of the Soviet system. The state dictated policy and ruled society; therefore it was assumed that anything to be learned about the Soviet experience would come from the study of the state. Lewin maintains that our long-term neglect of other social systems

[2] Peter Hauslohner, "Gorbachev's Social Contract," *Soviet Economy* 3 (January–March 1987), at 58.

[3] S. Frederick Starr, "The Soviet Union: A Civil Society," *Foreign Policy* (Spring 1988), at 27. Starr argues that "much of the initiative for change has shifted from the Communist Party to society." Because of the party's neglect or denial of social realities, Soviet society no longer perceives the Communist Party as its leader; since the party has disposed (intentionally or unintentionally) of many of its instruments of control (ideology, charismatic leader, terror), it can no longer force its subjects to follow.

in the Soviet Union has left us unprepared to understand the totality of the conceptualization of Gorbachev's reform efforts.

Lewin submits that the Soviet state (like all states, regardless of their degree of despotism) is itself susceptible to pressures and influences from outside factors. Even under Stalin, the state had to battle for its policies and could not always control and mold spontaneous movements in society to fit its policies. In the 1930s, for example, workers in industry jumped from job to job looking for better wages; the peasants who were pushed into kolkhozy exhibited an indifferent attitude toward work there; and bureaucrats covered up realities with which they thought the leadership would be displeased (pp. 25–26).

Lewin's central thesis (writing in 1988) is that the political structure of the Soviet Union had not changed to accommodate the vast shifts that have taken place in the Soviet social system. Urbanization, industrialization, and the growth of the professional and intellectual classes signaled a transition to a new historical era; now, Lewin argues, it is Gorbachev's task to reform all spheres of the system simultaneously. Indeed, Gorbachev's acknowledgment that these spheres are interconnected (in particular, his emphasis on the importance of the coincidence of economic and political reform) differentiates him from his predecessors.

During the last sixty years, the Soviet Union has been transformed from a predominantly peasant-rural society into a predominantly urban one. In 1940, 33 percent of the Soviet population lived in urban areas; by 1981 that figure had risen to 63 percent.[4] The mass urbanization of Soviet society involved a qualitative transformation, from relative homogeneity to considerable heterogeneity. Over the last several decades, the number of people willing to be employed in agriculture and positions of physical labor has been declining, while the number aspiring to nonmanual jobs in industry, information, and services has been increasing. The intelligentsia is growing relatively faster than the working class.[5]

Educational standards have improved dramatically during these decades. Secondary education is now compulsory for all children, and more

[4] Blair Ruble, "Social Dimensions of *Perestroyka*," *Soviet Economy* 3 (April–June 1987), at 176.

[5] Manual workers, of course, still comprise the largest group in the Soviet work force. Presently, approximately 58% of the work force are blue-collar workers (of these, 10–12% are unskilled or low skilled; 44–46% are skilled physical laborers; 3–4% are highly skilled), about 38% are officials and specialists, and 1–5% are *kolkhozniki* (Lewin, 51). Because the economy craves labor, the authorities are always trying to pressure young people to enter educational tracks that will lead to nonprofessional jobs. The problem is that the Soviet economy, stuck in the industrial stage of development, cannot accommodate the occupational aspirations of its young students.

people are entering into specialized secondary training and higher ec cation. In 1930, only 300,000 individuals received a higher education the Soviet Union. By 1960, the number had risen to 2.4 million, and 1980, it was 5.2 million. The increases in secondary specialized and pı fessional/technical school enrollment are similar. In 1930, 600,000 in(viduals were enrolled in secondary specialized education programs; 1960, that number was 2 million, and by 1980, it was 4.6 million. In 19: there were 250,000 students in professional/technical schools; by 19 there were 1.1 million, and in 1980, 4 million. (All figures are appro; mate.)[6]

Yet the Soviet economy is still "labor-hungry": there is what I Hewett has called a "major unemployment problem with machines, which appears to indicate a discrepancy between the educational level employees and the economic needs of the country. Periodic attempts reform in the Soviet education system over the last forty years are ev dence that the leadership itself acknowledges this discrepancy. In tl mid-1950s, Khrushchev began a program to combine education and pr(ductive labor in the classroom, to return the principle of polytechnic instruction to the secondary schools, and to discourage students froı pursuing study programs that would lead to higher education. The poir of the reform, according to Roy Medvedev, was to bridge the gulf th: divided school from later life, to satisfy the needs of the economy, and t correct the one-sided orientation of students toward higher education In spite of such attempts to steer young people into productive-labo programs, the number of individuals enrolling in higher education ha grown rapidly. Gorbachev is thus faced with the problem of making th economy fit the educational standards of the country, not vice versa.

Other important consequences of urbanization beyond more and bet ter education and occupational heterogeneity have put subtle but rea pressure on the state to rethink and revise its social policies. The value of the urbanized population, and especially of the better-educated an(professional elite, have undergone considerable change. In the formatioı of modern cities, people are often separated from family and friends anc must learn to survive outside the security of the collective that is so pre· dominant in a rural setting. To cope, they must learn to develop as in· dividuals primarily concerned with personal life and activity indepen· dent of any collective, including the state (Lewin, 63–66). Gail Lapidus

[6] Roger A. Clarke and Dubravko J. I. Matko, *Soviet Economic Facts 1917–1981* (New York St. Martin's Press, 1983), 44–46.

[7] Hewett, *Reforming the Soviet Economy. Equality versus Efficiency* (Washington, D.C., Brookings, 1988), 91.

[8] Medvedev, *Khrushchev* (New York: Anchor Press/Doubleday, 1984), 130.

has recognized several broad shifts in values that have marked the maturation of the educated elite: a shift from an emphasis on *partiinost'* (party élan) to a preoccupation with universal moral concerns; a shift away from collectivism to an increased recognition of individual and social diversity; connected with that, open recognition of conflicting interests and the admission that there can be no "single truth"; and a gradual retreat from a yearning for utopia to a growing realization that there are limits to "social engineering."[9]

These broad shifts in values, combined with the improvements in educational standards and the increasing relative importance of nonmanual and intellectual labor to the economy, are indicative of an increasingly independent, private, and self-sustaining broad stratum of Soviet urban life. People in this stratum are prepared and trained to contribute to the intensification of the economy, and are capable of participating in political processes. Yet the economy underutilizes them and the old social contract discriminates against them; indeed, that contract has until now deprived them of any real participation in the political life of their country. The question also arises: how many workers are now capable and willing to take on the same responsibilities, for the workers, too, have benefited from higher educational standards and other developments attributable to the urbanization of Soviet society.

In *Russia and the West*, Jerry Hough takes the thesis a controversial step further, arguing that the broad educated public of the "Gorbachev generation" is infected with a "Western orientation" due to a number of social processes that took place as this generation came of age in the 1950s and 1960s. While largely agreeing with the trends and shifts in values recognized by Lewin and Lapidus, Hough argues that much of the xenophobia and extreme censorship of Western ideas formerly prevalent in Soviet society reflected the values of the generation born at the turn of the century rather than the unchanging Russian national character or the inner laws of the Soviet system. He maintains that later generations have different values: they grew up in an era in which urbanization and industrialization were taken for granted, in which the war alliance with the United States and the West was an early childhood memory, in which patriotism and nationalism eroded old ideological dogmas, and in which the younger people in particular absorbed the Western values of the nineteenth-century *obshchestvo* (society) through Russian literature that had reappeared under Stalin.

Thus, not only the educated and trained stratum of Soviet society, but

[9] Gail Lapidus, "Gorbachev and the Reform of the Soviet System," *Daedalus* (Spring 1987), at 8–9.

many of the members of the political leadership itself, includi bachev, are eager to dismantle what Hough calls the "two Ir tains"—the one that was erected to keep out Western ideas and that was erected against Western market forces. They are willin card the old social contract which, by depriving the populatior opportunity to participate in the political process and the circul information, and by creating blue-collar economic egalitarianism, to maintain the Iron Curtains (p. 7). A new contract, Hough i would have to be closer to Western-style social contracts, which i many more aspects of the system, primarily civil and political as Such a contract would allow the leadership to tap broader huma sources more comprehensively.

Thus, the authors under review support the thesis that the Soviet tem has moved beyond the restricted Hobbesian contract, in which political leadership alone dictates its terms. The broad, educated e matured through the twin processes of industrialization and urbani tion and their by-products, is forcing a redefinition of the relations tween the state and its citizenry. In addition, as Hough reminds us, younger members of the political leadership have been manipulated the same social forces: many of them, including the president and g eral secretary, share a set of values, and consequently are part of the so force that is putting pressure on the first post-Stalin social contract.

It would be a mistake, however, to argue that individuals in the Sov system have consciously or intentionally influenced the process of sh ing the contract. Decisions regarding work, education, political outlo and so on are made by individuals for personal reasons, not by grou seeking to influence the social policies of the leadership. Participation the sense of Locke and Rousseau requires individual involvement guard the contract and to validate it continually. Multicandidate el tions to the soviets and congresses, in which non-communists can r and win against members of the Communist Party, are an important st in that direction. The balance of power over policy (and thus over t contract) is just beginning to shift. In the meantime, the population's pa participation in nonpolitical spheres of life has in some ways changed t reality in which the regime must make decisions.

Relationship between State and Society: Autonomy versus Class

In *Socialism's Dilemmas: State and Society in the Soviet Bloc*, Walt Connor examines the question of how, if at all, the interrelationship b tween state and society has changed in the Soviet Union. He describ

e Soviet state as having been born in isolation from social class (in 1917, e urban proletariat it supposedly represented was only a small percent- e of the population in Russia); so far, it has been able to maintain a gh degree of independence of any class. The party-state, not wanting become tied to particular social classes for its authority, has tried to sist the solidification of classes in Soviet society, classes that could even- ally reduce the state's autonomy. Connor asserts, however, that while Soviet society itself has not remained static over the last seventy years, the rapid changes that took place during industrialization are largely over, and the structure of that society now has a stable shape. Connor's case study of the working class serves as a basis for understanding the relationship between state and society.

The transition from agriculture to industry is mostly over, and many of the opportunities for upward mobility of the 1920s and 1930s are no longer available. As a result, Connor reports, the working class has be- come largely self-reproducing. He then raises the question whether Soviet workers consider themselves a "class"; that is, whether they are conscious that they have common interests and common goals, and that hey could therefore assert themselves collectively as a potential political and economic force. His answer is that there is no evidence that Soviet workers assert their rights and needs as a class—that, in fact, "mass" concerns remain within the borders of the workplace (job security, de- cent pay, availability of consumer goods, and the like), and that the po- iticization of the workers in the near future seems unlikely.

I would argue, however, that the strikes that broke out in the coal mines of the Kuznets Basin in the summer of 1989 show some weak- nesses in Connor's analysis because: (1) Although these strikes were pri- marily concerned with improvements in work conditions, the important point is that the strikers, by using the strike as a political tool (though not primarily for political ends), forced the government to bargain openly. (2) The strikers did achieve one overtly political goal—the re- placement of appointed trade union officials with their own men. (3) The list of grievances presented during the strikes in Siberia that took place in the fall of 1989 included calls for the repeal of Article 6 of the Consti- tution which, until March 1990, upheld the dominant position of the Communist Party in Soviet society. A discussion of whether these actions signify a hardening of the class consciousness of workers is beyond the scope of this paper; it seems clear, however, that the workers showed a certain level of political maturity by forcing the government to negotiate, rather than simply dictate, economic and political policy in the work- place.

ers. Furthermore, Soviet leaders are now unable or unwilling to regulate the flow of information from outside—especially in the form of Western books and video cassettes; and so Soviets have other standards of living and of knowledge with which to compare their own. Thus, the party does not control the population's economic and private behavior to the extent it once did. Through lack of discipline at work and attraction to outside knowledge and products, individuals are expressing their dissatisfaction with a social contract that does not reward them for hard work either in their paychecks or at store counters. Thus, although the party may be the only actor (in the sense of Locke and Rousseau) with final decision-making power in the area of social policy, other forces are pressuring the regime to validate its legitimacy through an update of the social contract, in ways partially defined by the nature of these forces themselves.

Job Security and the Welfare State

The state does have an interest in changing its social policy. After all, one of its primary concerns is to secure its legitimacy and stability, and to block forces that appear to be weakening it. Bialer (writing in 1980) maintains that the performance of its domestic economy is one of the chief sources of the stability of the Soviet regime.[14] In 1979, however, economic growth started to decline sharply; it stayed at a low level into the 1980s. Anders Aslund, in *Gorbachev's Struggle for Economic Reform*, argues that this decline, combined with the government's recognition of serious social problems such as alienation from and apathy toward the functioning of the economy (which are manifested in statistics on the theft of state property and drunkenness and in a weak work ethic), gave impetus to perestroika. Both Aslund and Lewin assert that rapid industrialization has left the economic structure struggling to keep up with the public's demands for a service-oriented economy. Aslund concludes that it was the realization that "significant improvements of Stalin's system were no longer feasible" that led Gorbachev to adopt a comprehensive approach to economic reform (p. 111). On the whole, Aslund's book lacks a conceptual framework of how the Soviet economy functions. It is, however, a useful summary, more descriptive than analytic, of Gorbachev's economic reforms between 1985 and 1988.

Aslund does make clear that the leadership is concerned about the economy's ability to deliver goods and services to the population. Because

[14] Bialer (fn. 10), 148–58.

Brezhnev actively pursued a policy of developing the economy's capacity to produce consumer durables, the standard of living of the average Soviet citizen increased during the 1960s and 1970s. The Brezhnev regime's commitment to investment in agriculture is well documented. Both meat consumption and living space per capita increased, as did the percentage of families owning refrigerators, washing machines, television sets, automobiles, and other consumer goods.[15] At the same time, the scope and extent of the state's welfare program improved. The size of retirement pensions increased, peasants became entitled to social security benefits, the minimum wage was raised, and so on. The Brezhnev regime had made a conscious decision to increase consumption relative to investment; as a result, the rate of economic growth continued to slow down.[16]

Another aspect of Brezhnev's economic policy was an unyielding commitment to the egalitarianism and security that had been established under Khrushchev as the basis of the first post-Stalin social contract. The Soviet government had issued to workers what Hewett has called "an insurance policy against personal economic risk."[17] Soviet-style socialism promises the right to work to everyone who is of age and able; full employment is possible largely because of an excess demand for labor, which, in turn, is the result of an extensive approach to growth and the inefficient utilization of labor. In the past, workers had an implicit understanding with the regime: retail prices for basic necessities would remain stable, and wages would continue to rise. Indeed, the monthly incomes of tens of millions of people increased dramatically between the mid-1960s and early 1980s. The average monthly salary (in 1982 rubles) for the state-employed labor force (that is, excluding kolkhozniki) jumped from 90 rubles in 1964 to 173 rubles in 1981.[18] Thus, the workers have always enjoyed job and income security, and have been protected from the economic uncertainties that persist in many of the Western capitalist democracies.

Economic egalitarianism was also an objective of the Brezhnev regime; it is a bias that has come to be taken for granted among workers and managers. Wages for those employed in the same type of job tend

[15] Meat consumption for the entire population increased from 38 kilograms per capita (per annum) in 1964 to 57 kilograms in 1975, but leveled off after 1975. In 1965, there were 22 television sets per 1,000 people in the Soviet Union; by 1980, there were 249 per 1,000. Refrigerators: 29 per 1,000 in 1965, 252 per 1,000 in 1980. Washing machines: 59 per 1,000 in 1965, 205 per 1,000 in 1980. See Clarke and Matko (fn. 6), 24–25, 28.

[16] It should be noted, however, that the growth rate of national income (national income utilized) decreased dramatically in the 1970s and 1980s. Hewett (fn. 7), 57.

[17] Ibid., 39.

[18] Clarke and Matko (fn. 6), 40–41. Although wages rose in this period, growth rates of productivity fell; therefore the buying power of a paycheck did not greatly increase.

to be equal, regardless of the individual worker's performance and the profitability of the enterprise. Losses incurred by individual enterprises are often covered by the profits of other enterprises, and are thus socialized. On a larger scale, when one sector does not perform well, the entire society pays for it. Aslund cites as an example the relationship between Soviet energy and Soviet agriculture. The export of energy is the country's foremost earner of hard currency; however, this currency is spent abroad for the purchase of grain because domestic agriculture does not produce enough to feed both the population and its livestock.[19] If it was not needed to make up agriculture's deficit, this hard currency would be used in other ways.

For most of the Soviet population, egalitarianism and economic security are gains that their system has won for them. They constitute the conspicuous achievements of the Soviet socialist economy and are a source of support for the regime among the population. Soviet workers have learned to expect artificially low housing rents, virtually free utilities, and subsidies on food and other consumer products as a part of the socialist welfare package (Connor, 134). These guarantees are so much a part of the post-Stalin contract that it is risky for the government to tamper with them. In the early 1960s, for example, riots broke out in the southern region of the Soviet Union after price increases on basic food items had been announced. (In Poland, such price increases have brought down more than one leader.) Brezhnev preferred to tolerate massive subsidies on food products to cover the gap between procurement and retail prices rather than to permit a decrease in the real incomes of the population. The fear of the potential consequences of violating this aspect of the compact seems to permeate the thinking of many economists, both conservative and reformist, in the Soviet Union; it explains why the Gorbachev regime avoided the issue for the first five years of its administration. Aslund quotes a Goskomtsen (State Committee on Prices) economist who argued for keeping retail prices low in order to ensure the "political stability" of the country (p. 131).

The costs of maintaining this aspect of the social contract have been very high, however. Hauslohner argues persuasively that economic efficiency and growth have been the victims of the guarantees of egalitarianism and security. Here scholars tend to get lost in a complex relationship of cause and effect that is not easily disentangled. The problem revolves, essentially, around economic modernization and the legitimation of the regime. According to Hauslohner, the two are "mutually con-

[19] Hewett cites a similar example (fn. 7), 48.

straining"; that is, improvement in the economic system is constrained by the state's legitimation strategy of security and equality; legitimation, in turn, is constrained by the performance of the economic system.[20] He maintains that overall economic growth and a bigger national income pie, plus improvements in the supply of consumer goods and services and a better social infrastructure, would be sources of support for the government. It is Gorbachev's task to break out of the cycle.

How do economic guarantees harm economic growth and consequently, in Hauslohner's schema, the state's legitimation? Hewett argues that the key incentives for getting workers and enterprises to produce efficiently and to innovate are stifled by the concept of "ceilings" and "floors," on which economic security rests, within the enterprise structure. Enterprises know that they will not be allowed to fail or go bankrupt; so there is a permanent safety net, or floor, beneath them.[21] Because profits resulting from efficiency or innovation are cross-subsidized to support poorly performing enterprises, a ceiling exists that automatically discourages enterprises from trying harder.[22] Aslund confirms that in this system of socialization of losses and profits, the "beneficial treatment of inefficient loss-making enterprises and the stifling severity towards the most profitable" effectively discourage any economic initiative.[23] There has been little room for ambition and risk taking in the Soviet economy.

As with the enterprises, so with the workers. While workers can be fairly confident they will not be laid off, they also know that hard work is rarely rewarded and that shoddy work is rarely penalized.[24] The incentive problem has bred a difficult cycle within itself. The general shortage of consumer goods gives the workers even less incentive to work hard, yet they must work harder to help alleviate the deficit in consumer goods. Ultimately, then, the economic guarantees that are implicit in the social contract are *disincentives*, and are at least partially responsible for the lack of labor discipline and for workers' indifference toward work. Furthermore, the costs of food imports, budget subsidies,

[20] Hauslohner (fn. 2), 54.

[21] There is still no law on bankruptcy per se, although the topic has been widely discussed in the press. The possibility of liquidating or selling unprofitable enterprises was mentioned in the Law on State Enterprises (1986), but little progress toward an actual bankruptcy law has been made.

[22] Hewett (fn. 7), 207–10.

[23] An enterprise's profits may be taxed up to 90%; the rate is determined by the ministry within whose jurisdiction the enterprise falls. Interest charges on capital range from 0 to 12%.

[24] It has recently become easier for enterprise directors to lay off redundant or unproductive workers; Connor points out that directors no longer have to find alternative jobs for workers before they can release them (p. 125). Presumably such workers must fend for themselves at local job placement centers in their search for new work. See also Aslund, 82.

and excessive wage increases—all undertaken in order to preserve the contract—have been overbearing.

Most important, however, is that Gorbachev's ideas on economic reform are blocked by the social contract. To reform the economy, he must reform the social contract. He must seek to legitimize the regime with different payoffs for the population. Instead of total job and income security, there must be more and better consumer goods on the store shelves; instead of general equality of wages, there must be faster national economic growth. This transition will not be easy. For more than thirty years, people have considered the economic guarantees as a quid pro quo in their socialist system. Now Gorbachev maintains that the old wage policy stifles productivity and must be changed. Personnel in nonmanual and service-sector positions have traditionally been underpaid relative to the ideologically glorified working class of manual laborers. During the Brezhnev era, wage policy discriminated against two particular groups: professionals and diligent workers. Because there was no distinction in pay between a "good" worker and a "bad" worker, the long-term effect has been to destroy the incentive to work. To correct this situation, Gorbachev has called for greater wage differentiation *among* groups in society, as well as *within* groups in society. Obviously, such a policy will discriminate against less productive workers (Connor, 126–27).

Another important step has been taken toward greater (and undoubtedly resented) differentiation of incomes. It is relatively easy to keep incomes leveled in a society in which the state owns all property, including the means of production. Under Gorbachev, however, the Soviet state has begun to loosen its grip on property. Cooperatives have been formed to provide consumer goods and services (restaurants, furniture, car repairs, and so forth). In the agricultural sector, private activity has been greatly expanded. The size of peasants' personal plots has been increased, along with the amount of livestock and poultry they can raise. They can build farm buildings and greenhouses with state credit, and they are allowed to form cooperatives. In March 1990, the Supreme Soviet of the Soviet Union passed a law that permits individuals and organizations to own land and to pass it on to their relatives. However, they do not have the right to buy and sell land for profit. Profits may only be made from working on the land—either by erecting a small business on it or by farming.[25] These measures indicate that the state will not be overly concerned if incomes differ widely among citizens in private and coopera-

[25] "O zemle," *Pravda*, March 7, 1990, pp. 2–3.

tive endeavors and those in the state sector. Fairly severe taxes on cooperative profits are being assessed, however, so that the state will benefit as well. Moreover, a highly progressive personal income tax was introduced in March 1988.

The utility of the policy of full employment has also come under severe attack recently. The threat of unemployment might provide a useful tool for disciplining labor and for forcing workers to increase their productivity. Many Soviet economists argue that enterprises should follow the same principles: if they are performing at their potential, they will stay in business; if they are not, they will "lose their job," that is, go bankrupt. The possibilities of unemployment and bankruptcy, like the possibilities of wage inequality and private economic activity, represent the Gorbachev regime's firm bias on the side of efficiency; Hewett aptly calls this the "equity-efficiency trade-off."[26]

How will labor, and especially lower-skilled and unskilled workers, react to the destruction of the economic aspect of the social contract and its replacement with greater uncertainty and market discipline into the workplace and economy? The new social contract undeniably places lower-skilled and unskilled workers at a disadvantage. Under the old policy, they were "winners," but under a redefined contract, they can quickly become "losers"—and many will. In the transition from an industrializing to an industrialized economy, many workers will become superfluous. During the stage of extensive development, labor is highly valued as an input, and there is excess demand for it. With the transition to an intensive approach, the productivity and skill of a labor unit becomes more valuable than purely massive quantities of input. Thus, poorly trained and unskilled workers would be the first to lose their jobs. The earlier implicit social contract had kept that from happening. In the transition to a more intensive approach to economic modernization, the same workers could easily be losers on the wage front, too, as skilled and trained labor will be more highly paid than unskilled and untrained labor. On top of that, as prices of basic goods rise, real incomes are likely to decline. (The leadership has promised, however, that wage increases will correspond to price increases.)

In order to restrict the number of workers who are unhappy with the prospect of a redefinition of the social contract, Gorbachev will have to produce some sort of payoff, in the form of more and better consumer goods and food products. In that case, all workers will be better off as consumers. The reworking of the contract—in which the workers did

[26] Hewett (fn. 7), 3.

not directly participate—will then hold some concrete benefits for them.

What happened to *gospriemka* (state control over the quality of products, created under Gorbachev) is a good example of the conflicting needs and goals of the citizen as worker and the citizen as consumer. Starting on January 1, 1987, *gospriemka* was introduced in selected enterprises. The idea was that checkers would reject low-quality goods, and that workers' pay would vary with the percentage of goods accepted. Consumers would benefit in the long run, but below-standard workers would obviously lose out. When the quality controllers did reject goods of poor quality (after all, they are frustrated shoppers, too), some workers lost their bonuses (Aslund, 76–80). Enthusiasm for *gospriemka* disappeared quite rapidly under social strains caused by disgruntled workers; according to Aslund, these strains became "intolerable" and "politically unacceptable" (p. 80).

Labor can thus be seen either as a constraint on Gorbachev's efforts toward economic reform or as a force that advances these efforts. Weak labor discipline, high rates of alcoholism and absenteeism (which rose rapidly after Gorbachev gave up his "dry" campaign), and low worker productivity may be considered either a result of guarantees of job security and basic necessities or a form of protest against the system's inability to provide acceptable recompense to the workers in the form of differentiated incomes and quality consumer goods. In the late 1970s, the performance of the Soviet economy worsened considerably; for the workers, this meant a continuing slide in the growth rates of real per capita income.[27] A slowdown in production is an expression of dissatisfaction, a response to the decline in the rate of improvement of living standards that is more subtle than (but just as real as) more obvious forms of protest, such as demonstrations and strikes. The growth of labor productivity decreased dramatically in the 1970s and 1980s in all spheres of the economy.[28] The regime must take the workers' revolt against shortages of consumer goods seriously before it can expect more efficient production from them.

The cooperatives upon which Gorbachev laid great hopes have not been particularly successful in easing the deficit of goods and services. (By contrast, the salaries of cooperative employees are generally higher than those of state employees. The result has been distrust of and animosity toward the cooperative movement.) The government is now hoping its economy will improve with the introduction of a "regulated mar-

[27] Ibid., 62–63.
[28] Ibid., 52.

ket economy," in which some types of private ownership would be encouraged, production and banking would be "demonopolized" (the government would maintain management of transportation, communications, energy, and defense), republics would be allowed much more economic freedom, retail prices would be reworked to reflect costs more realistically, and the ruble would eventually be made convertible.[29]

PROGRAMS OF INCLUSION: DEMOCRATIZATION AND GLASNOST

It is possible that Gorbachev will find a partial solution to the problem in a redefinition of the other aspects of the social contract. Democratization in the workplace is intended to serve as a payoff to the workers in the economic sphere, especially since it may be some time before more and better consumer goods start showing up in the stores. According to Hauslohner, workplace democratization—in particular elections of managers and broader workers' rights vis-à-vis management—has been presented as an integral part of the entire social policy package. It is, "in effect, a 'good' which the leadership is offering in exchange for other goods to be taken away, in the hope of diffusing whatever protest these take-backs may arouse."[30] Lapidus speculates that Gorbachev has concluded that a greater sense of participation will at least partially compensate the Soviet people for the absence of other immediate payoffs, in return for their participation in the reform effort.[31]

Several things should be pointed out here. First, it is relatively easy to democratize the workplace. This concept has been discussed for a long time. However, its political implications are not confined to the enterprises. Not only would the balance of power between workers and management be disturbed; more importantly, the control of local party functionaries over enterprises could be upset. Long used to interfering at will in an enterprise's affairs (and often to the detriment of the enterprise), these functionaries would be forced to abide by the decisions of elected enterprise representatives. The result would be a serious loss of power for them locally. This type of democratization, however, has had strong support among the leadership; steps have already been taken to democratize the workplace. The Party Plenum of June 1987 endorsed both the idea of enterprise councils (an institutional alternative to trade unions for representing workers' interests within the enterprise) and the idea of

[29] "K gumannomu demokraticheskomu sotsialismu: Programmnye zayavleniye XXIII s'ezda KPSS," *Pravda*, July 15, 1990, pp. 1–3.
[30] Hauslohner (fn. 2), 75.
[31] Lapidus (fn. 9), 22.

elections for managerial positions (including that of director) within the enterprises.

The coal miners' strikes of summer 1989 are evidence of the reality of the politicization of the working class *from below*. There has been much debate over whether Soviet workers perceive their needs and interests as a group, and whether they are capable of organizing and thus, of becoming a political force. The significance of these strikes does not lie in what they gained by their demands—rights that many other enterprises already had, such as the right to go on *khozraschet* (whereby an enterprise itself, rather than the ministry, covers expenses and reaps profits), higher pay for night shifts, Sundays off, and so forth.[32] Rather it lies in the fact that they compelled the government to bargain openly. The miners formed strike committees and set demands, and the government responded with its own committees; subsequently, compromises were reached.[33] This outcome was a first small but definite step toward the possibility of real interaction and communication between the Soviet government and Soviet society—which is a critical prerequisite for the establishment of civil society.[34] After the summer strikes, workers in the region formed the "Union of Kuznets Basin Workers" to defend their rights and to look after their economic well-being. The union organized a congress in spring 1990, which convened in late April, to coordinate independent workers' groups and movements from around the country. The workers began their own newspaper, *Nasha Gazeta*, in December 1989, and their representatives appear regularly on "*Pul's*," a local television program. Last April, a journalist writing on the situation in the Kuznets Basin complained in *Pravda* that *Nasha Gazeta* was attacking the CPSU and had, on at least one occasion, described socialism as "rotting."[35]

Second, Gorbachev seems to be stressing the importance of professionals in the economy and society, and to be paying less attention to blue-collar workers. That is quite a departure from fifty years of glorification of the working class and the emphasis on the worker as the key to building socialism. The new attitude is reflected in wage policy reform. In

[32] "Trebovaniya gornyakov udovletvorny," *Sotsialisticheskaya Industriya*, July 22, 1989, pp. 1–2.

[33] The strike committees in the Donets and Kuznets basins are still organized and active.

[34] The right to strike, which was pushed, ironically, by the official trade unions, was conceded by the government during the miners' strikes. However, many industries that are essential to the functioning of the national economy—including mining—are not allowed to strike. So far, the miners' strike committees have not been ordered to disband; they have declared that they will strike again if the compromises that were made are not adhered to.

[35] "Bespokoistvo: Tak mozhno opredelit' politicheskuyu situatsiyu v regione," *Pravda*, April 28, 1990, p. 3.

1985 and 1986, the government raised the salaries of scientists, researchers, designers, technicians, teachers, and professionals in the health services. In the summer of 1990, Prime Minister Ryzhkov promised further wage increases to people employed in what he called the "cultural sphere." Wage increases for blue-collar workers, on the other hand, are supposed to depend directly on whether their enterprise turns a profit. This means two things: first, they will literally have to work harder for an increase; second, workers' wages will differ at different enterprises, a potentially demoralizing factor.

The point is that we must look beyond the idea of workplace democratization as a payoff to broader movements for better and wider participation in the country's political processes, and beyond democratization as merely a support builder and protest blocker. The working class as a part of the citizenry with which the state must work out a contract has become relatively less important. The better-skilled are the natural constituency for Gorbachev's reforms. But, as pointed out above, Gorbachev must elicit at least the consent and the compliance of the working class. He does not want the lower-skilled and unskilled workers to become protestors *against* the reform effort. Because he cannot expect that they will give their active support, he is hoping for their compliance.

Gorbachev's program cannot proceed (much less succeed) without actual support from a broad segment of the Soviet population. A government can usually remain in place with just the compliance of its population. It is more difficult to fight the ingrained status quo than to fight movement against it. Compliance is not enough, however, when a government desires to change the status quo. Gorbachev's reforms require *active* support from a large part of the Soviet population. One of his greatest concerns now is that cynicism concerning his reform package and distrust of his leadership have recently grown within the Soviet Union. Although in 1987 a large portion of society was excited by the prospect of change, many became disinterested in the political situation in 1990. Passive compliance and apathy, even though they do not constitute dissent, can block the reform effort.

Soviet society has always been largely passive in its relation to the state. The party and the state led, and the people followed, although not always actively. It will be difficult to break this habit; as Brzezinski points out, communist parties in power have always found it easier to control societies than to mobilize them.[36] In the economic system, in official cultural expression, and in politics, the initiative and participation of the people

[36] Brzezinski (fn. 13), 167–68.

has been largely muted; in some cases, it has been nonexistent. The first post-Stalin social contract reflected the generally passive role of citizens' involvement in normal social functions. Although they had economic guarantees, they were completely excluded from the political realm. The power of the system, which ideally should reflect the dynamics and maturity of a society, was for the most part determined by the Communist Party and beyond the reach of the people.

This deliberate policy of exclusion has come to haunt the Soviet leadership. Two specific processes have been at work simultaneously for a long time. First, the composition and nature of the Soviet population has changed quite radically from when the party came to power over seventy years ago; the shift reflects a general (if not mass) process of maturation and civilization within society, manifested by trends in education and urbanization. Second, the party superstructure has ignored social realities by refusing to relinquish the vanguard role, which, undeniably, it had in the first years of Soviet power when it was more developed than the base it was leading. It has been unwilling to accept a follower role, in which it would become a reflection of predominant social forces, much like the parties in the advanced industrialized democracies. The result is that it has remained aloof from other social forces and that it has not benefited from the dynamics and the continual development of the society it is supposed to represent. In short, the party has not managed to tap the vast human resources that Soviet society offers. And it is at this point, when Gorbachev desperately needs to change the passivity of the status quo and to reform the system, that many of his generation are finally realizing the party's miscalculation.

As a result, Gorbachev has begun to implement reforms in the political sphere which, in essence, could redefine the social contract toward more active participation on the part of the Soviet public. Effective participation by that public in the country's political processes would provide a communications link between society and the party. With this redefined aspect of the social contract, the system would have the opportunity to develop, in Lewin's words, the "habit of dialogue" with its citizens. In seeking his constituency among professionals and trained workers, Gorbachev might strengthen their support for his reform package by broadening participatory rights. Hauslohner uses the term "middle-class consumption goods" in speaking of glasnost, cultural liberalization, hard news, and the like.[17]

In *Soviet Politics, Political Science and Reform*, Ronald J. Hill used the

[17] Hauslohner (fn. 2), 73.

writings of Soviet political scientists to study the debates on "democratizing" the Soviet system that have been going on within the Soviet Union since the 1950s.[38] The major practical issues of the past have been problems with the electoral system (including absence of candidate competition, nonsecret ballot, and inertia in voting), weakness of the representative system, underutilization of public opinion, lack of freedom of information, and—resulting from these inadequacies in the system—the underdeveloped political consciousness and culture of the masses. According to Connor, "Soviet political culture lacks a citizenry with participant orientations towards politics adequate to support movement . . . toward democracy" (p. 57). In the past, Soviet political culture has been parochial; that is, people were concerned with the daily affairs of surviving—procurement of food, pay raises, living conditions, and so on; they have been notoriously apolitical on a national level.

Political reform is a long-term process. Hill argues that the political culture of the Soviet Union has been a *subject* culture and that the first step of political reform is transforming it into a *participant* culture.[39] Much depends on the citizens' ability to act as participants in their political system. In order to make informed decisions and choices, people must have wide access to information and must be well educated in the workings of political processes in their country. Moreover, the system should provide institutionalized outlets for public opinion.

So far, the policy of glasnost has been stretched to surprising limits. The public has been exposed to debates incorporating many different viewpoints on almost all topics. Within Soviet society, the flow of information has become quite open, although the party still controls the major means of mass communication. Lenin is openly criticized, and capitalism is openly advocated during the frequent panel discussions broadcast on Soviet television. The home-produced two-page newspaper *Lumpen*, which advertises itself as an "unofficial anticommunist publication," is sold freely on Moscow's streets, as are the liberal newspapers from the Baltic states and other nonconformist "mini-newspapers." (One such paper attracts customers with its sensationalist title, *Tainy Kremlevskogo Dvora* [Secrets of the Kremlin Yard]. Last winter *Pravda* published the complete programs of the former Democratic and Marxist platforms of the CPSU, although it severely criticized both and in general led the campaign against factionalism in its pages.)

The nature and extent of Gorbachev's higher-level political reforms

[38] Hill, *Soviet Politics, Political Science and Reform* (Oxford: Martin Robertson, 1980).
[39] Ibid., 12–13.

have already gone farther than many in the West expected. In March 1989, the first Congress of People's Deputies was elected directly by the public in multicandidate, secret-ballot elections. The Congress is comprised of representatives from three groups: organizations, territories, and nationalities. (The first of these had to be officially approved; therefore, the "Democratic Union," for example, could not have its candidates elected through the "organizations" channel, but the Komsomol, on the other hand, could.) The Congress elects the Supreme Soviet, which in turn has been granted legislative powers. The Congress has also formed many oversight committees to watch over certain interests, such as defense, transportation, and so on. It was these committees that vetoed several of Prime Minister Ryzhkov's choices for ministry heads in summer 1989. Popularly elected congresses and soviets exist for all the Soviet republics and their localities.

For the time being, Gorbachev, as head of the executive branch of the government, seems to have solid control over the Congress, which is run by his party colleague Anatoly Lukyanov. At present, about 85 percent of the deputies are members of the Communist Party. Close to 700 of the 2,250 deputies have only a high school or agricultural education. Perhaps not surprisingly in a system that in the past has had little respect for the law, only forty-five deputies have any legal training.[40] Gorbachev is thus working with a Congress that is not prepared to make too many waves on national issues, although there are certainly individual deputies who are quite vocal in disagreeing with the regime's policies. Some debates concerning social issues such as pensions and taxes have been quite bitter.

At the national level, therefore, no serious opposition has yet arisen that could challenge the Communist Party for seats. Gorbachev has obviously had more trouble keeping control over the congresses in the republics. For example, even in the Congress of the Russian Republic, in which more than 86 percent of the deputies are party members,[41] Gorbachev's candidate for president of the republic was rejected last spring, and Boris Yeltsin was elected instead. Furthermore, in the local soviets across the country, an average of only 49 percent of the elected deputies are members of the Communist Party.[42]

On the whole, there has been definite movement toward involving the public more in the political sphere, and consequently toward reducing

[40] *Narodnye Deputaty SSSR* (fn. 1).

[41] "Deputatskiy Korpus Rossii," *Pravda*, March 26, 1990, p. 2. Interview with Vitaliy Vorotnikov, then Politburo member and Chairman of the Presidium of the Supreme Soviet of the RSFSR.

[42] Ibid.

the barrier that separates the party from society. The redefinition of the political aspect of the social contract will help Gorbachev's reform program in two distinct ways. It may help to build broader support for that program (as people realize they have a stake in it)—especially if the reform ends up forcing the Communist Party to compete on equal terms with other parties. Also, just as more information will be flowing from top to bottom, more information will be flowing from bottom to top; the party will benefit from this more open relationship with society.

Conclusion

This essay has developed three themes: (1) Gorbachev has acknowledged that the post-Stalin contract seriously hampers his efforts toward reform. (2) The communist regime still monopolizes the country's national-level political posts as well as the Congress of the Soviet Union. In making final decisions on social policy, however, it is influenced by various groups in society, as information on their responses to existing social policy (that is, what they choose to study, where they decide to work, whether they subscribe to communist ideology, and so forth) flows upward to the decision makers. (3) The forces pushing for a new social contract appear to be firmly entrenched; they face resistance, however, from groups whose jobs and benefits are no longer secure, such as blue-collar workers and party apparatchiki.

So far Gorbachev's program has mainly encompassed policy changes *within* the social contract. "Goods" are being traded—in essence a more efficient economy for more active participation in support of the regime. But the *method* used to maintain the contract over time has not yet been altered, even though the institutionalization of legislative and executive branches of government and the repeal of Article 6 are certainly first steps in this direction. In theory, the government could eventually be controlled by another group. Rousseau proposed that individuals must continually consent to the contract for it to be valid; the General Will was to be determined through a system of voting. The process by which Soviet citizens could consent to preserve the contract has begun to be developed, but for the time being the Communist Party maintains a real monopoly on policy, even though the Constitution no longer guarantees that monopoly.

Gorbachev seems to be telling Soviet society that although he is more appreciative of its evolution than his predecessors were, as head of party and state, he can still make decisions on policy as he sees fit. The national-level Congress by no means has complete control over all legis-

lative activities. On the other hand, Gorbachev, intentionally or not, has allowed competition to be created in local areas. Many important questions have still, however, not been answered. Within Soviet politics, will the balance of power (and power over the contract) shift away from the Communist Party? Will the party allow itself to be seriously challenged on the national level as it is being challenged on the local level? And if it is challenged successfully, will the party relinquish its power? Just how far Gorbachev's original theses of glasnost and democratization will go toward presenting a real challenge to the regime's power in society—that is, toward creating a legitimate multiparty system under equal conditions—should be seen in the near future.

RATIONAL DETERRENCE IN AN IMPERFECT WORLD

By BARRY NALEBUFF*

I. Introduction

THIS paper applies recent economic research on games with incomplete information to the calculus of deterrence.[1] We do not attempt to provide a complete theory of rational deterrence. The focus is on the role of reputation and signaling in establishing deterrence. We show how strategic players signal a reputation for strength from the perspective of an internally consistent equilibrium model.

The theory of rational deterrence is based on the application of cost-benefit analysis to conflict initiation. The costs of conflict are compared with the benefits of cooperation; if all parties prefer cooperation, then the status quo will prevail and a potential conflict is avoided. Even at that level of generality, the theory is incomplete. A critical missing ingredient is the role of expectations.

A country's valuation of initiating a conflict depends on its belief about an adversary's intentions. This belief in turn depends on the perception of the adversary's own cost-benefit calculations. The fundamental problem is that all of these calculations and expectations are based on imperfect and incomplete information. The costs and benefits calculations combine subjective and objective elements. Parties in conflict are likely to view the situation from very different perspectives. How does rational deterrence apply when there is no objective set of calculations on which all parties can agree? In this paper, we show how to proceed with the necessary cost-benefit calculations in the presence of incomplete or imperfect information. The method is based on looking for a set of self-confirming beliefs. Consequently, there is no longer any guarantee that the calculations will provide a unique answer. In spite of this indeterminacy, we believe that the result is a much richer theory of rational deterrence.

The fundamental conceptual difficulty is the necessity of forming be-

* This paper has benefited from the generous comments of Michael Intriligator, Robert Jervis, John Vickers, and my SOM Thursday lunch colleagues. I thank the Pew Charitable Trust for financial support through their grant to the Princeton Center of International Studies.

[1] In truth, the calculus of deterrence might better have been called algebra.

liefs in the event that something which should never happen happens. For example, imagine that the cost-benefit calculation indicates that deterrence should work; then what do we believe about an adversary if deterrence fails? It would seem that a failure of deterrence indicates a failure of the model, and hence using the model to form expectations would seem to be a futile exercise. This is false. One must imagine the counterfactual possibility in order to perform the cost-benefit comparison that shows why deterrence should succeed.

The predicted probabilities derived from a rational deterrence model are endogenous and thus fundamentally different from the probability associated with the meteorologist's forecast of rain. If the weatherperson predicts a 100 percent chance of rain and it fails to rain, this should serve as conclusive evidence that the forecaster's model is flawed. But if a model predicts that deterrence should be 100 percent successful, that prediction can only be made by considering the costs and benefits associated with the event that deterrence fails. Our procedure for calculating the endogenous probabilities considers the "impossible" event in order to demonstrate that it will never happen.

This paper has two objectives. One is to defang the reputation paradox of Jervis. The second is to explain and motivate recent refinements in equilibrium theory. Each of the theoretical results is illustrated with a simple arithmetic example. The resulting equilibria offer different perspectives on how to interpret potentially misleading reputations and suggest several resolutions to the reputation paradox.

Section II begins with a general discussion of reputation. The possibility that reputations can be strategically manipulated leads to a paradox. This paradox is described in Section III, using the *Mayagüez* rescue as a case study. One resolution to the reputation paradox is presented in Section IV, where we present the idea of a sequential equilibrium. Increasingly sophisticated methods of interpretation lead to the refinements of Section V. Section VI offers a brief conclusion.

II. Reputation

A critical part of any rational deterrence calculation is to input the relevant payoffs. What should we believe about others and what should they believe about us? If preferences and payoffs were all known, then payoffs would be objective and specification straightforward. But in the presence of uncertainty and with the possibility of misperception, the payoff inputs are no longer clear. There arises a need to signal our objectives and to interpret the signals of others. How we communicate our objectives

or how we hide them) becomes an integral part of the rational deterence calculations.

One primary component of communication is the use of reputation. No cost-benefit calculation can be done in isolation. Each incident beomes part of the history for the future. Consequently, the cost-benefit alculation that leads to certain behavior in the present reflects on how ne will act in future cases. During the Berlin crisis in 1961 John F. Kennedy explained the importance of the U.S. reputation: "If we do not neet our commitments to Berlin, where will we later stand? If we are ot true to our word there, all that we have achieved in collective secuity, which relies on these words, will mean nothing."[2]

The problem is that everyone has an incentive to talk tough. Making he correct inference is not straightforward. In a chapter entitled "Signals nd Lies," Robert Jervis explains: "Whether the state is lying or honest t will try to issue those signals which it thinks will get its desired image ccepted. . . . Both an honest man and a liar will answer affirmatively if sked whether they will tell the truth."[3]

Thus we judge actors by their actions rather than their words. Reputation, based on a long and consistent history of behavior, helps predict he future. The Mayflower furniture company along the Massachusetts Turnpike proudly advertises that it has gone 127 years without a sale. Are they still waiting for their first customer?) This unconditional comnitment to everyday low prices brings in a year-round steady stream of ustomers. A sale might temporarily raise profits, but it would be anther 127 years before they could repeat such a clever advertisement. Next year, we expect the sign will read 128 years. The reputation beomes self-perpetuating as it becomes more valuable.[4]

Rarely do politicians or countries have the opportunity to establish uch a long-standing reputation. Moreover, the application of the reputation is not nearly so much in their control. The reputation of the Maylower furniture company is sufficient to deter one particular event, a ale. Countries are not so fortunate; reputations must be built to deter nforeseen future events.

A country's reputation is multidimensional, and yet there are rarely nore than a few observations on which to base beliefs. The problem with ntegrating reputation into the cost-benefit calculations is that one must xtrapolate out of the sample. How might President Kennedy's response

[2] Quoted in F. Ikle, *How Nations Negotiate* (New York: Harper and Row, 1964), 67.

[3] Jervis, *The Logic of Images in International Relations* (Princeton: Princeton University ress, 1970).

[4] Sadly, we must report that the Mayflower furniture store recently had its first sale, a oing-out-of-business sale.

to Berlin forecast the U.S. response to a Soviet invasion of Europe? Did Kennedy's reputation extend to President Johnson? Could the U.S. invasion of Panama be predicted from the response in Grenada?

The problem is further complicated by the endogenous nature of the inferences. *The value of a reputation depends on how others interpret it.* This in turn affects the willingness to create the reputation in the first place. The strategic use of reputation cannot be evaluated in isolation. It must be part of an equilibrium model.

The advantage of a formal model is that it forces us to integrate the consistency requirements of equilibrium with the strategic use of signaling. Reputation is presented as a continuous variable. Consequently, the issue is no longer whether to believe but *how much* to believe based on what one sees. The equilibrium level of reputation defines the extent to which one may attempt to extrapolate the future from the past.

III. Signaling and Deterrence: A Paradox

To discuss the role of signaling and reputation as part of deterrence, we start with the work of Robert Jervis. His analysis has been at center stage beginning with *The Logic of Images in International Relations*. An interesting and seemingly counterintuitive theme from Jervis is that signaling may be counterproductive. A country may get stuck in a negative feedback cycle. He uses the *Mayagüez* rescue as a representative case study.

After the U.S. retreated from Vietnam, it was important for the Americans to reestablish a reputation for toughness. The rescue of the *Mayagüez* seemed to provide the perfect opportunity. On Monday morning, May 12, 1975, a U.S. vessel called the *Mayagüez* was captured by a Cambodian torpedo boat. The Cambodians accused the ship of spying.[5] President Ford first attempted to negotiate the release of ship and crew. But there was no diplomatic response from the Cambodians. Instead, they fired at the U.S. reconnaissance planes flying over the *Mayagüez*. On Wednesday evening President Ford sent in the marines. Although there were heavy marine casualties (partly due to poor intelligence about the ship's location), the captain and his thirty-eight crewmen were rescued. The *New York Times* reported as follows:

> [Both politicians and the public] saw the event as a re-assertion of American will after this country's disorderly retreat from Indochina. . . . Ever since the American evacuation from Saigon, Administration officials had

[5] But after an interpreter was found, the crew did not appear to be in immediate danger. The Cambodian captors were apparently persuaded that the *Mayagüez* was not a spy boat.

been saying frankly that America's international stature could be restored by a demonstration of strength.[6]

While the stated purpose was solely to rescue the *Mayagüez* crew, even ıe headlines read "Ford Sends a Signal." The *New York Times* article ›ntinued: "The Administration has been specific about one nation, [orth Korea, to which it would like to send a clear, strong signal since ıe defeat in Indochina. Now they say the signal has been sent: Don't ıake a move against South Korea without expecting American military ıtervention."

Was it rational to expect that rescuing the *Mayagüez* would restore the [.S. reputation for strength? If so, then the *Mayagüez* rescue would have een done even by a "weak" U.S. government, for this would be a small rice to reestablish its reputation. But then observers should realize that ıtervention is no longer a signal of strength. Nothing is learned from ıe American behavior, since both a tough and a weak U.S. would purıe the same strategy. Jervis explains the difficulty in drawing inferences:

> This raises [a] problem with such inferences about resolve. Beliefs like these will undermine rather than support themselves: if the U.S. felt that fighting a small war for a country of little intrinsic value would lead others to conclude that it would display high resolve in a dangerous confrontation, then this action would not provide reliable evidence because the U.S. would fight the small war in order to create a favorable impression irrespective of whether or not it would run high risks in a nuclear exchange.[7]

If we take this argument to its logical conclusion, the U.S. allies (and lversaries) should pay no attention to the U.S. theatrics of acting tough. has no signaling value. This brings us around full circle. If acting ›ugh does not improve one's reputation, then what is the reason for a 'eak party to pretend to be tough? The negative feedback seems to dispate any possible improvement in reputation.

Improvement is, of course, relative. While the reputation may not imrove relative to the status quo, the status quo is no longer an option. 'hus we care about the relative effect of intervening on reputation as ›mpared with not intervening. A country that fails to act could suffer a ıassive loss in reputation. Avoiding the loss is what keeps the cost-:nefit calculations positive.

The problem with this interpretation is that there is a lack of discipline ı determining what one can and cannot believe. In order to determine

[6] *New York Times*, Week in Review, May 18, 1975.
[7] Jervis, "The Symbolic Nature of Nuclear Politics" (The Edmund James Lecture, Deırtment of Political Science, University of Illinois at Urbana-Champaign, 1985).

when a party should intervene, we must simultaneously determine when it is best not to intervene. Neither calculation can be done in isolation.

The intuition behind the formalization of equilibrium is that expectations are in fact negatively equilibrating. The more desirable a certain action is, the less one infers about those who take the action. There is less of an effect on reputation and consequently there are fewer circumstances in which the action is taken. This helps restore the implications about those who act. In equilibrium, the effect on reputation is just sufficient to motivate those who act and no others.

Note that an essential part of this story is that reputations are continuous. There is the possibility of drawing greater or lesser inferences. It is this move away from black and white beliefs that allows us to find an equilibrium in spite of the negative feedback effect. To see the problem with black and white beliefs, we return to an analysis from Jervis. Read it with an eye toward calculating what one should believe, not what one should not. Jervis writes:

> An ironic possibility should be noted. A concern for reputation can lead states to act and draw inferences in a pattern opposite from the one that we—and most other analysts—imply. This is not to dispute the common starting point; states refuse to back down not because of the immediate and direct costs of doing so, but because of the belief that a retreat will be seen as an indication of general weakness and so lead others to expect retreats in the future. But the desire to counteract such undesired consequences may lead a state that has retreated on one issue to pay especially high costs to avoid defeat on the next one. Thus the United States was not only willing but anxious to use force to free the *Mayagüez* because it wanted to show others that its evacuation of Indochina didn't mean it would not defend its other interests—the very consequence which it had predicted would follow from a defeat in Vietnam and which had justified its participation in the war. If others understand this logic and expect states to behave in this way—to follow retreats with displays of firmness—then reputations for carrying out threats do not influence estimates of credibility because—to compound the paradox—reputations are so important that states must rebuild them when they are damaged. If you have been caught bluffing in poker, others are likely to call you in the next round in the belief that you bluff a lot or are they unlikely to do so because they think you know it is no longer safe to bluff? To the extent that the latter is the case, perceptions of credibility are influenced by the state's recent behavior, but in a way which produces equilibrating negative feedback rather than the positive feedback of domino dynamics.[8]

The poker story shows the seeming paradox of self-falsifying inferences. If one believes that a player who recently bluffed will now play

[8] Jervis, "Deterrence and Perception," in Steven Miller, ed., *Strategy and Nuclear Deterrence* (Princeton: Princeton University Press, 1984), 66–67.

afe in order to reestablish a reputation, then there is an incentive to ontinue bluffing. But if this is the belief, then others will continue to all and the player should play his cards straight. Neither inference can e an equilibrium. What is?

One option left out is the possibility of a randomized (or mixed-trategy) solution. The need for a mixed strategy is really a consequence f the artificial assumption that there are only two types of actors, weak nd strong. To restore continuity, a reputation is based on a probabilistic elief about an actor's type—a reputation is the inferred probability that he actor is strong.

Imagine that a strong country always acts tough. The reputation effect rom acting tough will depend on what a weak country is expected to lo. If it was thought that a weak country would never attempt to bluff r act tough, then seeing a country act tough would indicate true trength. This would improve its reputation immensely, possibly enough ven to motivate a weak country to act tough (which would not be an quilibrium). As bluffing becomes increasingly likely, the enhancement f a reputation following tough behavior is diminished. At some point, he probability of bluffing is sufficiently high (and the improvement in eputation is sufficiently small) that the cost of the weak country acting ough is exactly offset by the gain in reputation. This is the mixed-trategy equilibrium.[9]

Although the mixed-strategy approach provides a consistent solution, may have a somewhat artificial flavor. This is solely a consequence of ne highly stylized assumption that there are only two types of actors, eak and strong. The problem is both simpler and more realistic when e allow for a continuous range of incomplete information.

[9] Although we have found a consistent solution, is it of any use? Here Jervis has some oubts as to the ability of game theorists to say much more about the "black box" of mixed rategies:

> Economists have not been able to model the behavior of oligopolists nearly as deterministically as they have that of the wheat farmer facing a market he cannot influence. . . . [I]n many situations game theory prescribes a mixed or randomized solution. Of course, game theory yields great insights into how actors try to out-think and out-bluff each other, *but the competitive and variable sum nature of the situation means that scholars cannot produce deductions on the model*: "In situation X the actor always should (or will) do Y." We can show that the actors' calculations are consistent with deterrence, but this requires looking into the "black-box" of decision-making. (Jervis, "Rational Deterrence: Theory and Evidence," *World Politics* 41 [January 1989], 183–207)

n fact, it is possible to characterize the precise equilibrium proportion of mixing. Or to put more poetically, there is a method to the madness of a random strategy; see Avinash Dixit nd Barry Nalebuff, *Thinking Strategically: A Competitive Edge in Business, Politics, and Ev-ryday Life* (New York: W. W. Norton, 1991). To verify whether or not actors follow their rescribed mixed-strategy rules does not require a large number of independent observations om a repeated game. Instead the theory is tested by evaluating how well it predicts cross-ectionally, looking across a variety of different conflict situations.

IV. Equilibrium

There is another way to tell the *Mayagüez* story. Instead of the artificial nature of just two types, weak and strong, imagine that the propensity to intervene comes in continuous varieties. To be more concrete, we consider a stylized model of conflict initiation.

There is an event that provides the U.S. with a payoff of x if it intervenes. The variable x should be thought to include all the observable elements of the cost and benefit calculations. In addition to the observable payoffs, there is an unobservable component, c. The variable c can be thought to represent all the psychic and other intangible costs and benefits associated with intervention. For ease of exposition, x is described as a benefit and c as a cost.[10] Thus, a low value of c indicates a greater propensity to intervene, as the unobserved costs are low. Conversely, a high value of c is evidence against intervention. Intermediate values of c indicate intermediate willingness to intervene.

The role for a reputation arises because countries do not know each other's c. Each side starts with some expectation about the distribution of the other's parameter. For analytic convenience, we take the initial beliefs to be uniformly distributed between zero and one.[11] These beliefs are updated based on observed behavior. Thus a country may act strategically in order to manipulate how others perceive its intervention cost.

Without any notion of reputation, the U.S. would intervene if $x > c$ and do nothing otherwise. However, because the U.S. cares what others think about its unobserved parameter, the effect on reputation enters into its cost-benefit calculations. Once again for convenience, let the value of a reputation equal $a\,[1 - \bar{c}]$. Here $\bar{c}$ is the average value others think the c is for the U.S. and a is a parameter that measures the importance of reputation. One interpretation of the reputation effect is that the unobserved costs and benefits are correlated over time and across different circumstances. Reputation captures the future value of changing others' perception of your unobserved costs.

Initially, $\bar{c} = 1/2$. When $a > 0$, the U.S. wants others to believe that its value of c is low since this will make them less likely to act against the U.S. interest. The parameters a and $\bar{c}$ provide a shorthand, or reduced form, representation for the value of a reputation. A larger value of a corresponds to a greater significance placed on reputation. A change in $\bar{c}$ corresponds to a revised belief about the country's cost of intervention.

[10] But there is nothing that prevents x and c from being negative, in which case the observed value would represent a net cost and the unobserved parameter would represent a net benefit.

[11] Here, the assumption that intervention costs are uniformly distributed is made solely for analytic convenience. The prior belief should be based on the history up to this point.

To model the outcome, we begin with a listing of the minimal re-
quirements for an equilibrium. Denote the set of types who intervene by
and those who choose not to intervene by N. There is an expectation
ibout the representative type of country that chooses to intervene; this is
lenoted by c_i. For a country that chooses not to intervene, the expectation
of its cost is c_n.[12]

1. *Maximizing behavior*. (a) The payoff to any country that intervenes is higher than if it does not intervene; (b) the payoff to any country that does not intervene is higher than if it does intervene.

2. *Consistency of beliefs*. (a) If the sets I and N are both nonempty, then the expectation of the costs of those in each set should be based on the distribution of types in each set; (b) if one set is empty (say N) and the other contains everyone (say I), then expectations about those in I must equal the prior belief, but there is freedom to form an expectation about who might be in N should this zero-probability event arise.

Together these requirements are called a sequential equilibrium, a
efinement of Nash equilibrium due to Kreps and Wilson.[13] It is impor-
ant to emphasize that these two conditions are minimal requirements
or an equilibrium. Maximizing behavior is an essential element of ratio-
iality. It imposes cost-benefit analysis as the basis for deciding whether
or not to intervene. The consistency of beliefs condition is more subtle.
The second part of this condition provides a degree of freedom that often
permits a spectrum of equilibrium outcomes. Consequently, more than
one equilibrium outcome may satisfy maximizing behavior and consis-
ency of beliefs. Some are more appealing than others. To choose be-
ween equilibrium outcomes, one may look to impose a stronger test of
ationality.

If, as part 2(a) supposes, we expect countries with one range of costs
o intervene and those with some other range not to, then our conclusion
ibout who did what should be based on the ranges corresponding to the
observed action (or inaction). But what are we to believe about the set of
ioninterveners if the expectation is that everyone will intervene? In case
2(b) we are forced to have an expectation, c_n, over who might be in this
iull set. Sequential equilibrium is agnostic in this matter and allows the
expectation to take any possible value. (For this expectation to be part of
in equilibrium, the reputation effect must then motivate everyone to
hoose intervention as their maximizing behavior.) It is possible to form
nore sophisticated expectations about who might have taken an action

[12] Note that we do not require that $c_i \leq c_n$. It is possible that intervening hurts one's reputation. Such an example is presented below.

[13] David Kreps and Robert Wilson, "Sequential Equilibria," *Econometrica* 52 (July 1982), 863–95.

that never should have happened. In Section V, we provide three increas ingly restrictive assumptions about how to form rational expectations i case 2(b). The three refinements of sequential equilibrium we conside are the successive elimination of dominated strategies, universal divinity and perfect sequential equilibrium. Each is defined and illustrated usin the reference example. We begin with the possibilities for a sequentia equilibrium.

There is an event that gives the U.S. a chance to improve (or worsen its reputation. Intervention is observed to be worth an amount x. To fin a consistent set of beliefs, we undertake the cost-benefit analysis assum ing there is some expectation about the unobserved cost for those wh intervene and those who do not. These expectations lead to a reputatio c_i for a country that intervenes and a reputation c_n for a country tha maintains the status quo. With these reputations in mind, the U.S chooses its optimizing behavior. The resulting behavior is an equilibriun if and only if the optimization confirms the initial expectations about th expected intervention cost for the U.S. when it intervenes and when i does not.

Were the U.S. to intervene when its true cost was c, its payoff woul be

$$x - c + a[1 - c_i].$$

The U.S. earns an observed payoff of x, pays an unobserved amount c i intervention costs, and ends up with a reputation valued at $a[1 - c_i]$.

If the U.S. does not intervene when its true cost is c, its payoff is ther

$$a[1 - c_n].$$

Note that this valuation is independent of x and c since x is not earnec and c is not paid; instead, the U.S. ends up with a reputation valued a $a[1 - c_n]$.

Hence, it is better to intervene provided that

$$c < x + a[c_n - c_i].$$

We call this critical cost c^*,

$$c^* = x + a[c_n - c_i].$$

As seems intuitive, when intervention costs are low $[c \leq c^*]$, the U.S acts, whereas when intervention costs are high $[c > c^*]$, the U.S. does no act.[14]

The calculations above provide a formula that reveals the critical valu

[14] How we decide the action for the c^* type is irrelevant.

* for any imagined values of c_i and c_n. Consistency requires that we xpect an intervener to have costs in the range [0, c^*] and that we base ur belief c_i on the prior distribution of costs in that interval. Similarly, oninterveners have costs in the range (c^*, 1], and so the expected value f c_n is based on the prior distribution of costs in this upper range. It is mportant to note that this consistency condition falls under 2(a) only when $0 < c^* < 1$. The two other possibilities are discussed in depth be-ow.[15]

Because the prior distribution is uniform, when intervention is ob-erved, the expected value c_i taken over the range [0, c^*] must be

$$c_i = c^*/2.$$

f no intervention takes place, then the expected value for c_n is

$$c_n = [1 + c^*]/2.$$

We have two equations and two unknowns. Any solution is a sequen-ial equilibrium. Taking the difference between the two equations shows hat for any solution, $[c_n - c_i] = 1/2$.[16] We first solve for c^* and then c_i nd c_n.[17]

$$c^* = x + a/2, c_i = (x + a/2)/2, c_n = (1 + x + a/2)/2.$$

To provide a numerical illustration, let $a = 1/2$, and consider an event with $x = 1/4$. Then $c^* = 1/2$, $c_i = 1/4$, and $c_n = 3/4$. There is a sequen-tial equilibrium in which all the types with intervention cost below 1/2 act whereas those with costs above 1/2 do nothing. Moreover, ruling out for the moment solutions with zero-probability events, this equilibrium is unique.

As the parameter values change, the predicted outcome varies in a natural way. A rise in the value of reputation, a, results in greater partic-pation. A higher reward, x, from intervening also encourages greater ntervention in equilibrium.

This simple model illustrates the cost of maintaining a reputation. The only way to prove something about yourself is to take an action that would be too costly if you are not who you are trying to convince others

[15] If $c^* = 0$ then no one should intervene, whereas if $c^* = 1$, then everyone should inter-vene. Both endpoint cases predict that one action is never taken and hence falls under the weaker consistency condition 2(b).

[16] This simple relationship is an artifact of the uniform distribution of intervention costs.

[17] This type of well-behaved sequential equilibrium exists provided $0 < c^* = x + a/2 < 1$. Otherwise, the effect of reputation may be so large or the value of intervening so great that we expect everyone to act, $c^* \geq 1$. Alternatively, the value of reputation may be so small (or negative) or the price of intervention so costly that all types choose not to intervene and $c^* \leq 0$.

you are. But unlike the circular reasoning of Jervis, this can be done in a matter of degree. The unobserved cost of intervention separates the weak from the very strong—not perfectly but some part of the way. Seeing a country intervene does not reveal its exact cost, c; it just sends a signal that the intervention costs are below rather than above c^* and thus average c_i rather than c_n.

An implication of the model is that we see too much intervention. Here too much intervention means that intervention takes place for its effect on reputation and would otherwise not be desirable. An intervener's gain in reputation is a nonintervener's loss. Put together the two effects exactly cancel out, and we have a zero-sum transfer. This excess intervention would not occur if the country's unobserved parameter were known. A country would intervene only if $x > c$. There would be no reputation effect since if c were known, its perception could not be changed.

When costs can only be inferred rather than observed directly, then the type with $c = x$ strictly prefers to intervene (and by continuity so will types with $c > x$). The reason is that intervention has little or no direct cost and results in a positive gain in reputation for being perceived as a c_i rather than a c_n.

The reputation effect is a two-edged sword. As Jervis emphasizes, a country that does not act cannot presume its status quo reputation remains unchanged. That effect is captured in the cost-benefit analysis. A country that fails to intervene sees the value of its reputation fall, not necessarily from the status quo but from the value to which it would have risen had it intervened. In our example, the status quo value of a reputation is 1/2. But once the opportunity to intervene arises, those who act are viewed as having an expected cost $c_i = 1/4$ while those who do not intervene are represented by an average cost $c_n = 3/4$—there is no option to remain at $\bar{c} = 1/2$.

Self-fulfilling Equilibria

So far the calculations have been straightforward. Now we turn to the more subtle possibility of an equilibrium based on self-fulfilling expectations. In these cases everyone is expected to choose the same action, either intervention or nonintervention. No one dares to be different because the loss of reputation associated with taking the "wrong" action dominates the direct gain or loss.

The possibility of multiple equilibrium based on different expectation foundations is described in Jervis: "There is a great deal of room for false

onsensus effects—i.e., if each person thinks everyone else holds a certain iew, then that view becomes the operating reality. . . . The truth of the atement depends on whether people believe it."[18]

This is the problem we now confront: to what extent can expectations ad to an equilibrium? What are the limits to which beliefs are self-ılfilling? It is here that the power of equilibrium reasoning has its great-st effect. We are able to show exactly what type of beliefs are internally onsistent. The scope for expectations to drive the model is remarkably mited.

The nature of these self-fulfilling equilibria relies on what might be alled a "lemmings" effect: it must be that *everyone* is expected to do the ıme thing in equilibrium.[19] There are two candidate equilibria. In one, veryone intervenes because the loss of reputation from not doing so verwhelms other considerations. The more types that act, the more ostly it becomes not to act. The expectations feed upon themselves until ven the weakest type is forced to intervene in order not to be exposed. 'his is just the limiting case of our previous example.

The other possibility is that no one intervenes; the reason is that an ıtervener is thought to be weak—not strong—so the reputation loss 'rom intervening outweighs any direct gain. This reversal of a reputation ·ffect is quite different from what we have previously discussed. To for-nalize these possibilities it is appropriate to return to the lurking issue of what to believe when we see something that we thought was not pos-ible.

One interpretation upon observing a zero-probability event is that the nodel is wrong. Since we have seen something that obviously should ıever have happened, we must have misunderstood the original situa-ion. This negative perspective misses the point of cost-benefit analysis. t is simply impossible to predict that something should not happen with-ut analyzing the payoffs if it did. The choice of action is based on a alculation of net gain. How can we choose an optimal decision knowing nly one side of the equation?

Rather than discuss this issue in the abstract, the model provides a evealing backdrop. Imagine that $c^* = x + a/2 \geqslant 1$. In this case we redict reputation is so important that everyone should be willing to ntervene. But how is this justified? It is clear what we should assume if ve see an intervention; since we expected everyone to behave this way,

[18] Jervis (fn. 7).

[19] Otherwise, there are no zero-probability events. The earlier analysis applies and there is ı unique solution.

it should have no effect on our prior beliefs. Yet this calculation was conditioned on the cost-benefit calculation, which by necessity must place a value on not intervening.

We must specify how others will interpret a failure to intervene. Since nonintervention is a zero-probability event, we cannot apply the standard technique to calculate posterior beliefs. The standard technique used to calculate posterior beliefs about who did what is Bayes' rule. The formula for the posterior probability that someone has cost c conditional upon observing nonintervention is

$$\text{Prob}[\text{cost} = c \mid \text{nonintervention}] = \text{Prob}[\text{cost} = c \ \& \ \text{nonintervention}] / \text{Prob}[\text{nonintervention}].$$

The problem is that nonintervention is supposed to be a zero-probability event so that both the numerator and the denominator are zero. Bayes' rule makes no prediction in this case. The question of what is reasonable to believe in this surprise event is an unresolved research question in game theory. The present model helps illustrate some of the possibilities. Here it is important to emphasize that there is more than one way to form rational beliefs and that reasonable people differ over which to choose.

One solution to this paradoxical situation is provided in the definition 2(b) of sequential equilibrium. Choose any beliefs over the range $[0, 1]$ for the types who do not intervene. The only constraint is that these beliefs must then be consistent with the initial assumption that no one would in fact want to intervene. The way to check if this is possible is to assign "pessimistic" beliefs in the event of this zero-probability outcome. For example, if the country fails to intervene, we believe that its unobserved costs are 1, the worst possibility. Given this belief, when is the model internally consistent? The payoff to intervention is $x - c + a/2$, while the payoff to nonintervention is $a(1 - c_n) = 0$. All types prefer to intervene when

$$x - c + a/2 \geqslant 0, 0 \leqslant c \leqslant 1.$$

This is true if and only if $x + a/2 \geqslant 1$.

Thus our intuitive belief that all types will intervene when $x + a/2 \geqslant 1$ is confirmed. When the observed payoff for intervention is large (compared with the unobserved costs) and reputation is highly valued, then it is natural to suppose that there is an equilibrium in which all countries intervene. Except for the issue of forming expectations in a zero-probability event, there is nothing different or unusual about this self-fulfilling equilibrium. The outcome is simply the limiting case of our previous example for $c^* \geqslant 1$.

The power of sequential equilibrium is that it restricts the extent to nich expectations can drive the outcome. Believing in something can-ıt always make it happen. In our numerical example, there is *no* se-ıential equilibrium that involves intervention by all parties. The reason that with $x = 1/4$ and $a = 1/2$, $x + a/2 < 1$; a country with interven-ın cost c close to 1 would prefer to have it presumed that $c = 1$ than lose nearly 3/4 in order to improve its reputation to $c_i = 1/2$.

HE REVERSED EQUILIBRIUM

'e now turn to the paradoxical case. It is often possible to find a second t of beliefs consistent with the requirements of a sequential equilib-ım. For certain parameter values there exists a sequential equilibrium which no intervention occurs. This solution is fundamentally different character from our previous examples. The out-of-equilibrium beliefs e reversed. We assume that a country that intervenes is *weak*, not 'ong. Restraint is the sign of strength. The interpretation is that a coun-y that intervenes is so much trying to prove itself that it reveals its true eakness; countries with low unobserved costs are sufficiently confident their capabilities that they are willing to forgo this intervention. Of urse, this interpretation is particularly convenient for a weak country, hich by doing nothing can pretend that it is so strong that it does not ıve to intervene.

Is it possible that these beliefs are consistent? To support an equilib-ım where intervention is supposed to be a zero-probability event, we sign posterior beliefs that an intervener has $c_i = 1$. The unobserved sts assigned to an intervener are as large as possible. The payoff to a untry if it intervenes is then

$$x - c + a(1 - c_i) = x - c.$$

Expectations following the case of nonintervention are more straight-rward. The presumption is that no one should intervene. Hence, ob-rving that a country does not intervene should have no effect on its timated intervention cost. A country that does not intervene has its putation preserved at the prior level, $c_n = 1/2$. The payoff for this 'ategy is then $a[1 - c_n] = a/2$. Not intervening is the best strategy for country with unobserved cost c when

$$a/2 \geq x - c.$$

nce the presumption is that no one intervenes, this inequality must be ue even for the country with the greatest propensity to intervene, the pe with $c = 0$. This implies

$$x - a/2 \leq 0.$$

Again, consider our numerical example with $a = 1/2$ and $x = 1/4$. In this case, there is a sequential equilibrium in which no type intervenes. The reason is that the loss of reputation costs 1/4, which is just enough to offset the value of intervening for all types.[20]

The advantage of the formal modeling is to put a consistency constraint on what might be thought of as an equilibrium. The examples illustrate that sequential equilibrium does not go very far in that direction. The solution concept seems to allow opposite extremes just by changing the interpretation of a zero-probability event.

V. Refinements

Sequential equilibrium is just the first step on the road to forming beliefs out of equilibrium. It is a minimal set of requirements. Consequently, there may be a multiplicity of equilibria. In this section, we introduce three refinements of sequential equilibrium. These refinements involve more sophisticated reasoning about interpretation of actions, especially those off the equilibrium.

The first refinement is based on the iterated elimination of dominated strategies; this is closely connected with the idea of rationalizability.[21] Alternative restrictions on beliefs are provided by universal divinity[22] and perfect sequential equilibrium.[23] These three alternatives to sequential equilibrium are discussed in turn.

One can judge the merits of these approaches in either of two ways: by their axioms or by their results. A comparison of the axioms shows three plausible restrictions on what might constitute a reasonable belief. As we move from elimination of dominated strategies to universal divinity to perfect sequential equilibrium, the restrictions employ increasing levels of sophisticated logic. A comparison of the results reflects this ranking. Each increased level of sophisticated reasoning further reduces the scope for self-fulfilling expectations to support a no-intervention equilibrium. Therefore, if one intuitively feels that this equilibrium is

[20] Note that if $a = 1/2$ and $x > 1/4$, then the model predicts that a country with $c = 0$ would strictly prefer to intervene. Hence the reversed equilibrium is ruled out.

[21] David Pearce, "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica* 52 (July 1984), 1029–50; B. Douglas Bernheim, "Rationalizable Strategic Behavior," *Econometrica* 52 (July 1984), 1007–28.

[22] Jeffrey Banks and Joel Sobel, "Equilibrium Selection in Signaling Games," *Econometrica* 55 (May 1987), 647–62.

[23] Sanford Grossman and Motty N. Perry, "Perfect Sequential Equilibria," *Journal of Economic Theory* 39 (June 1986), 97–119.

unreasonable, then the refinements provide a more general framework for characterizing just what makes it unreasonable. But one may well take the other view. There are circumstances where the no-intervention equilibrium seems reasonable (such as when there are no objective benefits from intervention) and yet it is ruled out. From this second perspective, the results present a specific example that questions the merit of the axioms. This paper does not attempt to take sides. The different approaches are presented in a way that leaves it to the readers to form their own opinions.

Elimination of Dominated Strategies

A strategy is dominated if there is some other strategy that leads to a higher payoff no matter how the action is interpreted.[24] The elimination of dominated strategies makes the presumption that a player will not follow a dominated strategy. In the present context this says that it is unreasonable to believe that a country should ever intervene if that action lowers its payoff (relative to not intervening) given *any* possible expectation about the unobserved cost parameter following intervention. And conversely, it is unreasonable to believe that a country should ever fail to intervene if nonintervention yields a lower payoff given *any* possible expectation about the unobserved cost parameter following nonintervention. In both cases, we eliminate dominated strategies from consideration.

To form beliefs about who will do what, we restrict attention to countries with a range of unobserved costs for which the proposed action is not dominated (if such a cost exists). The procedure is then iterated as illustrated below. With each cycle our beliefs are increasingly restricted as to who may take any action. Sometimes the beliefs will converge on a unique solution, thus selecting one of the sequential equilibria. More generally, we are able to show that some of the earlier reversed sequential equilibria are no longer sustainable with this restricted set of beliefs.

What are the possible beliefs? For the action that everyone takes, the posterior belief must equal the prior belief ($c = 1/2$). The worst belief about a country that fails to follow the proposed equilibrium behavior is that its $c = 1$. Hence, the changed perception of c following intervention is at most $1/2$ and the best possible net payoff to intervening is then

$$x - c + a/2.$$

If this is negative, intervention is dominated by not intervening. In our numerical example, $x = 1/4$ and $a = 1/2$. Thus intervention is domi-

[24] Note that this domination is over other strategies, not over an opponent.

nated for all countries with $c > 1/2$. No matter what others think about the reputation of noninterveners, it is not worthwhile to intervene whenever $c > 1/2$.

This simple idea allows us to eliminate the (reversed) sequential equilibrium that results in no-intervention. The reason is that we may no longer suppose that an intervener has cost 1. A country with cost 1 finds intervention to be a dominated strategy. The worst reputation we can assign to an intervener is to believe that its cost is 1/2—this is the highest cost for which intervention is not strictly dominated.

To see whether an equilibrium in which no one intervenes is still possible, we must also consider the other side of the cost-benefit ledger. What is the most favorable reputation that can be given to a nonintervener? Because intervention is a dominated strategy, all countries with costs from $(1/2, 1]$ must be counted as noninterveners. Therefore, the best reputation we can give to a nonintervener is to add all the types $[0, 1/2)$ to the list of noninterveners. This gives an expectation of 1/2 for those who do not intervene. These beliefs are illustrated in Figure 1.

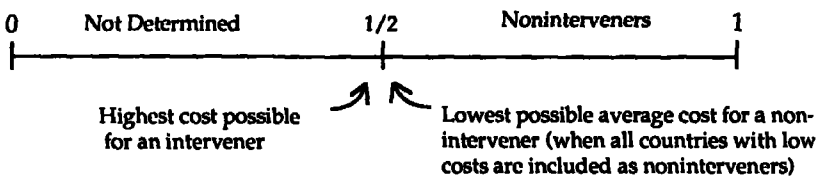



FIGURE 1

To illustrate the iterated elimination of dominated strategies, we now redo the cost-benefit calculations. The worst we can believe about an intervener is that its cost is 1/2; the best we can believe about a nonintervener is also 1/2. Thus when reputations are chosen to be maximally biased against intervention, we find that interveners and noninterveners are given identical reputations. The chosen action does not affect reputation. Hence all countries with $c < x$, or costs in the range $[0, 1/4)$ in our numerical example, will find intervention a dominant strategy.

This iterated elimination of dominated strategies limits our ability to use self-fulfilling expectations to punish those who intervene and reward those who do not. We must now expect that all those with costs below 1/4 will intervene while those with costs above 1/2 will not. As illustrated in Figure 2, our only freedom is to choose beliefs about those in the interval $[1/4, 1/2]$.

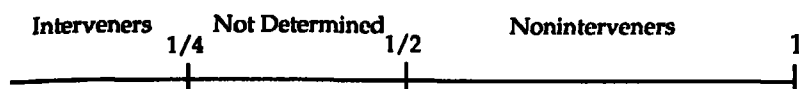



FIGURE 2

The worst reputation for an intervener arises when we presume that .l countries with unobserved costs from [0, 1/2] will intervene. The best :putation from the nonintervener's perspective follows from the belief ıat all types from [1/4, 1] will not intervene. (Note that these beliefs are ot internally consistent.) Reapplying the cost-benefit calculations shows ıat the reputation effect is now worth 3/8 so that countries with $c <$ '16 intervene whereas those with $c > 1/2$ still find intervention a domi- ated strategy. The range of ambiguity is reduced even further, down to ıe countries with costs in the interval [7/16, 1/2].

How we continue from here gets more complicated. But the limit ıould also be apparent. There is only one equilibrium that is left after 'e iteratively eliminate dominated strategies. The equilibrium is the one ı which countries with costs less than 1/2 intervene and those with costs ɔove 1/2 do not. This is the predicted outcome from the original sepa- ıting equilibrium as $c^* = x + a/2 = 1/2$.

The fact that elimination of dominated strategies leads to a unique ʒuilibrium makes that outcome particularly appealing. But even if the rocess did not converge to a unique solution, the usefulness had already :en demonstrated: for the case with $x = 1/4$ and $a = 1/2$, the reversed ʒuilibrium in which no one intervenes is not consistent with the iterated imination of dominated strategies. Although a weak country may aim that its restraint is a sign of strength, this is not a reasonable inter- retation given the parameters above. Here, a reasonable interpretation ıeans that we exclude the possibility that a country follows a dominated rategy. This restriction on what is reasonable to believe is an effective ɔol for limiting the scope of self-fulfilling expectations.

NIVERSAL DIVINITY

stronger restriction on beliefs is called universal divinity.[25] Universal vinity focuses on the consistency condition 2(b). Under a sequential ʝuilibrium one is free to assign an arbitrary reputation to a country that kes an action that should never occur. But is it reasonable to believe .at a country with high unobserved intervention costs would make an

[25] Banks and Sobel (fn. 22).

unexpected intervention while one with low unobserved intervention costs would not? Universal divinity classifies this as an unreasonable belief. We should not assume that a country that loses a lot by taking the unexpected action has deviated while those who would lose less do not. Instead, the argument is made that the expectation of who might have taken the out-of-equilibrium action should include all countries with negative payoffs whose losses are less than some cutoff amount. This idea has power to break a (reversed) sequential equilibrium in which no one is supposed to intervene; it has no power to break a sequential equilibrium in which everyone is supposed to intervene. We look at these two self-fulfilling equilibria in turn.

In an equilibrium where no country should intervene, the cost-benefit calculations must indicate a negative payoff to any country that intervenes.[26] Yet upon seeing an intervention, which country with a negative payoff is supposed to have taken this unprofitable action? Universal divinity argues that it is unreasonable to assume that those who stand to gain the least from intervention will do so without also believing that everyone else will deviate. Specifically, it is unreasonable to suppose that the intervener is a country with payoff -0.8 but not a country with payoff -0.4. One must choose some lower endpoint and take the average over countries with net payoff no worse than that amount. In our model this is equivalent to believing that the intervener has unobserved costs less than some critical c. The worst reputation for an intervener is to suppose that the critical c is as high as possible; everyone intervenes from cost $c = 1$ on downward. In this case, $c_i = 1/2$. This is also the reputation given to a nonintervener, $c_n = 1/2$. Since no one is supposed to intervene, a nonintervener's reputation must remain unchanged from its prior level of 1/2. Thus intervention does not improve reputation in this best-case scenario for nonintervention.

For the example with $x = 1/4$ and $a = 1/2$ and beliefs that accord with universal divinity, we cannot support an equilibrium where no country intervenes. A country with zero (or even close to zero) intervention costs is willing to intervene provided this does not hurt its reputation too much. Since we are not allowed to suppose that interveners have higher costs than noninterveners, there is at worst no loss in reputation. Hence, the cost-benefit calculation become strictly positive in favor of intervention for low-cost countries ($c < 1/4$). Just as in the case of iterated elimination of dominated strategies, we have broken the sequential equilibrium in which nonintervention is viewed as a sign of strength.

[26] Otherwise, intervention must be expected.

Universal divinity has no power to disrupt the sequential equilibrium n which everyone is presumed to intervene. In this case, the cost-benefit alculations must come out negative against those who might fail to in- ervene. The country for which this comparison is the least negative is he one with the highest intervention cost ($c = 1$). Hence, universal di- inity allows us to count as reasonable the belief that a nonintervener has $= 1$. There is no one else who finds nonintervention relatively more avorable and who thus must be added to the list. This implies a loss of eputation from $1/2$ down to 1 for a country that fails to intervene. If eputation were important enough (a is large) then all countries would hoose to intervene. For the cases with $c^* = x + a/2 \geqslant 1$, there remains self-fulfilling equilibrium in which everyone intervenes.

'ERFECT SEQUENTIAL EQUILIBRIUM

\ third approach is called perfect sequential equilibrium[27] or credible eologisms.[28] The idea is to look for an internally consistent set of beliefs bout who might have deviated. Is there a range of costs such that if we elieve that countries with those costs deviate then it is in their interest o do so and it is not in the interest of any of the other countries to leviate? If such an internally consistent belief exists, then the original equilibrium fails this third refinement test.

An example illustrates the technique. Consider a situation in which x s negative. Since it is common knowledge that there is no objective ben- efit from intervention, the preferable outcome is that no one intervenes. For the case with $x = -1$ and $a = 3$, there is such an equilibrium. The no-intervention equilibrium is supported by $c_n = 1/2$ and $c_i = 1/2$. Since here is no gain in reputation and intervention has an observed negative payoff, there is no reason for any country to intervene.

This solution satisfies elimination of dominated strategies. The best possible payoff for intervention is $x - c + a/2 = 1/2 - c$. Thus all countries with $c > 1/2$ find intervention a dominated strategy. This im- plies that the lowest value of c_n is $1/2$ and the highest value of c_i is also $/2$ (see Figure 1). But these are exactly the reputations used to support he prediction that no one should intervene. Unlike the earlier example, he group of countries for which $x > c$ is the null set and so there is no country that must now be classified as an intervener. There is no scope or iteration and the no-intervention equilibrium remains.

[27] Grossman and Perry (fn. 23).
[28] Joseph Farrell, "Credible Neologisms in Games of Communication" (Mimeo, University f California, Berkeley, 1984).

This solution also satisfies universal divinity. The requirement concerning beliefs about the zero-probability event is that we include all countries whose payoffs for that action are no worse than some amount. When intervention is the zero-probability event, then the lowest reputation arises when everyone is included (since the countries with low unobserved intervention costs must be included whenever high-cost countries are included). This leads to $c_n = c_i = 1/2$, which are the numbers used to support the no-intervention solution.

The point of this exercise is to show that the no-intervention outcome does not satisfy perfect sequential equilibrium (PSE). Whether this leads one to think the no-intervention outcome is less reasonable or PSE is less reasonable is left for the reader to decide. Under PSE, a country that observes the zero-probability event of an intervention must try to justify its occurrence. It is possible to do so using the following beliefs. One imagines that countries with unobserved costs from 0 to 1/2 will intervene ($c_i = 1/4$) and those with costs from 1/2 to 1 will not ($c_n = 3/4$). With these beliefs, the reputation gain from intervention equals 1/2, which is valued at 3/2. Comparing the unobserved cost of intervention with the observed payoff, we find that the beliefs are confirmed:

$$c^* = x + a/2 = -1 + 3/2 = 1/2.$$

Countries with $c < 1/2$ prefer to intervene and those with $c > 1/2$ prefer not to.

To put this example back into the definition of PSE, we have found a range of costs $[0, 1/2)$ such that if we believe that countries with those costs deviate (by intervening) then it is in their interest to do so and it is not in the interest of any of the other countries to deviate. Since such an internally consistent belief exists, the original no-intervention equilibrium fails the PSE test.[29]

Upon reflection, one might choose to argue that this separating equilibrium is the more sensible outcome. It is the outcome predicted by the base-case model for sequential equilibrium. Even though the average country ends up worse off, the countries with low unobserved costs benefit at the expense of those with high costs. The reputation of interveners is better by 1/2, and with the large weight placed on reputation this justifies paying the observable intervention costs, x.[30]

[29] Perfect sequential equilibrium is not a panacea. It is much harder to show that something satisfies PSE than to show that it does not. In fact, one of its failings is that it is possible that no equilibrium will satisfy this test. In our example this problem does not arise as the separating equilibrium satisfies PSE.

[30] As this intuition suggests, it is not the case that PSE eliminates all no-intervention equilibrium. If x is negative and a is small, the no-intervention equilibrium satisfies PSE.

This separating outcome may seem even more appealing if we examne the intuition for why the no-intervention equilibrium fails PSE. The roblem with expecting no one to intervene is that if we do observe the nexpected, this "mistake" can be justifed if we presume that the interener was a low-cost country ($c < 1/2$). But if we are willing to give this eputational advantage to a country that breaks the equilibrium by interening, for the sake of consistency we should penalize those who fail to ntervene ($c > 1/2$). The advantage of these expectations is that they offer consistent explanation for the observed behavior.

VI. Conclusions

The potential for a reputation paradox arises in a world with imperfect nformation. Because unobserved motivations matter in predicting beavior, there is a need to make presumptions about the unobserved costs nd benefits of any action. This is particularly difficult when it is an ction that cost-benefit calculations predict should never occur. The enlogenous relationship between the presumptions and the predictions omplicates the inference problem. The predictions are based on the preumptions, but the presumptions must be consistent with the predictions. When the predictions suggest that something should not occur, how are he presumptions to be made consistent?

Game theory offers (at least) four possible answers. We use the setting f a rational deterrence model to explain and motivate these recent reinements in equilibrium theory. The refinements illustrate what is pardoxical about an equilibrium in which no country intervenes because ntervention is associated with weakness rather than strength. The most ophisticated refinement can explain away the no-intervention outcome ven when it seems the more natural case, such as when the intervention ayoffs are all negative. The advantage of our stylized model is that it is ossible to follow the connection from assumption to conclusion. Here ne may choose among refinements by seeing the results. As with Christnas pudding, the proof is in the eating.

GLOBAL COMMUNICATIONS AND NATIONAL POWER
Life on the Pareto Frontier

By STEPHEN D. KRASNER*

Introduction

THERE is no single international regime for global communications. Radio and television broadcasting, electromagnetic spectrum allocation, telecommunications (telephone, telegraph, communications satellites, transborder data flows), and remote sensing are governed by a variety of principles, norms, rules, and decision-making procedures—or, in some cases, by no regime at all. Variation in outcomes can be explained by the interests and relative power capabilities of the actors in each case.

Global communications have been characterized not by Nash equilibria that are Pareto suboptimal but rather by disagreements over which point along the Pareto frontier should be chosen, that is, by distributional conflicts rather than by market failure. Changes in the relative power of states have led to changes in international regimes. The apparatus of economics, which has been so heavily deployed in regime analysis, has focused on information and monitoring rather than power, implying, if not explicitly arguing, that intelligence (figuring out the right institutional structure) is more important than the underlying distribution of capabilities. Regime analyses based upon market failure inevitably obscure issues of power because, given a Pareto suboptimal situation and a concern with absolute not just relative gains, it is possible to make at least one actor better off without making others worse off—an outcome that can be resolved through cleverness rather than by resort to power, threat, and coercion.

Information flows and knowledge have been less important than relative power capabilities for international communications regimes or the lack thereof. Where there have been disagreements about basic principles

* I would like to thank Peter Cowhey, John Ferejohn, Robert Jervis, Robert Keohane, Terry Moe, and especially William Drake, who is also especially absolved of any responsibility for the conclusions of this paper.

and where the distribution of power has been highly asym-
:ernational regimes have not developed. Stronger states have
: what they pleased. Radio broadcasting and remote sensing
arest examples.

ıere are coordination problems and the distribution of power
ıore symmetrical, however, regimes have been established.
conflict has varied according to whether states were dealing
ordination problems or with coordination problems that had
al consequences. The resolution of the former has caused lit-
ecause the purpose of the regime has been to avoid mutually
outcomes. The allocation of the radio spectrum before the
ıe telecommunications regime before the 1980s offer two ex-

ıst, in cases that have had distributional consequences, con-
n more intense: though the states agreed on mutually unde-
mes, they disagreed on their preferred outcome. Controver-
iggered by changes in power, usually resulting from the
t of new technologies. In recent years distributional ques-
recipitated conflict over the allocation of the radio spectrum
ternational telecommunications. The outcome of these dis-
en determined primarily by the relative bargaining power of
volved. Whereas previous institutional choices had not im-
constraint, new interests and power capabilities conferred
nologies have led to new institutional arrangements.

t to say that institutional arrangements were ever irrelevant:
were necessary to resolve coordination problems and to es-
lity. Without regimes all parties would have been worse off.
ıowever, many points along the Pareto frontier: the nature
nal arrangements is better explained by the distribution of
ver capabilities than by efforts to solve problems of market

Explaining International Regimes

ıences of different configurations of interests for the creation
ıance of international regimes has been elaborated in a num-
:s over the last decade. There are four possible configurations
two of which can give rise to international regimes. Regimes
nt for situations characterized by either harmony or pure
conflict. Rather, regimes can arise under what Arthur Stein

has called dilemmas of common aversions or dilemmas of common i
terests.[1]

Under zero-sum conditions there is no basis for regimes and no reas
to coordinate policies, because one actor's loss is another's gain.[2] In
situation of harmony, too, there is no reason to create a regime, becau
each individual player, acting without regard for the behavior of othe
maximizes both its own utility and that of the system as a whole. Pur
self-regarding behavior produces both a Nash and Pareto optimal eq
librium.[3]

By contrast, dilemmas of common aversions and dilemmas of co
mon interests are distributions of preferences that do create incentives
establish and maintain international regimes. Both involve strategic i
teraction. Dilemmas of common aversions refer to situations in whi
actors must coordinate their policies by agreeing on some set of rules
conventions, to avoid mutually undesirable outcomes. The specific co
tent of these rules will matter only if the actors disagree about which
the most desirable outcome. If there is no disagreement, then the ou
come is a Nash equilibrium and is Pareto optimal: there is no incenti
for any actor to defect and no opportunity to increase any actor's utili
without damaging that of another. Cheating is therefore not a proble
There is no need to develop elaborate mechanisms for generating a
monitoring information, because there is no sucker's payoff to wor
about. One set of rules is as good as any other, provided that all stat
agree to do the same thing. For instance, starting from a situation
which no prior investments have been made, it does not matter wheth
cars drive on the left or right side of the road provided that all drive
adopt the same rule. It may not matter to a couple whether they go
the mountains or the ocean for a vacation, provided that they go to t
same place. Such a configuration of interests is shown in Figure 1.

With a minimal level of coordination (actors need only avoid switc
ing back and forth at the same time when one starts from the right a

[1] Stein, "Coordination and Collaboration: Regimes in an Anarchic World," in S. D. Kr
ner, ed., *International Regimes* (Ithaca, N.Y.: Cornell University Press, 1983). Duncan Sni
refers to this same distinction as "Coordination versus Prisoners' Dilemma: Implications f
International Cooperation and Regimes," *American Political Science Review* 79 (Decemb
1985), 923–42.

[2] The essence of conventional realist thinking as exemplified in the work of Kenne
Waltz is that the quest for power, which is inherently a relative concept, inevitably pla
states in a zero-sum situation. See Waltz, *Theory of International Relations* (Reading, Ma
Addison-Wesley, 1979), as well as Joseph Grieco, "Anarchy and the Limits of Cooperatio
A Realist Critique of the Newest Liberal Institutionalism," *International Organization*
(Summer 1988), 485–507.

[3] Robert O. Keohane, *After Hegemony: Cooperation and Discord in the World Political Eco
omy* (Princeton: Princeton University Press, 1984), 51.

	Mountains Left	Ocean Right
Mountains Left	1, 1	0, 0
Ocean Right	0, 0	1, 1

FIGURE 1
SIMPLE COORDINATION PROBLEM

ıe other from the left), all actors will end up driving on either the left · the right, or both members of a couple will go to either the mountains · the ocean.[4]

Dilemmas of common aversions, however, can involve questions of ıstribution in addition to those of coordination. Although actors may ·cognize that all would be worse off without some agreement, they may isagree about precisely what the terms of that agreement ought to be. 'here are many points along the Pareto frontier.

Such a distribution of interests is shown in Figure 2. This game is ›mmonly referred to as the Battle of the Sexes. The story line here is as ›llows: Both members of a couple prefer to do something together, but ıey disagree on their preferred outcome, vacationing in the mountains r at the ocean. With such a distribution of interests, the choice of moun-ins or ocean, or left or right, matters, for vacationing in the mountains ves the column player a higher payoff than going to the ocean and vice :rsa. Both parties are averse to an absence of coordination in which they

	Mountains Left	Ocean Right
Mountains Left	2, 3	0, 0
Ocean Right	0, 0	3, 2

FIGURE 2
BATTLE OF THE SEXES

[4] I am indebted to Joanne Gowa and Susan Woodward for suggesting what I hope is a nder neutral example. As Snidal (fn. 1), 931, notes: "Sometimes coordination is presented nply as the problem of two or more actors matching policies where they are indifferent out where they match. . . . Here there is no disjuncture between individual and collective ionality and no problem of collective action. It requires not more than communication and nmon sense to achieve an outcome that is both individually and collectively optimal."

take different vacations, but the payoff matrix itself provides no information about which of the two equilibrium points will be chosen. The problem is not how to get to the Pareto frontier but which point along the frontier will be chosen.

In the international relations literature that has evolved since Schelling, commitment has been the analytic device that has most commonly been used to understand which outcome will be chosen, and commitment has usually referred to cleverness (such as throwing a steering wheel out the window in a game of Chicken).[5] The resolution of distributional conflicts could, however, be resolved through a very different route: the exercise of state power, which could be manifest in at least three ways.

1. Power may be used to determine who can play the game in the first place. In international relations less powerful actors are often never invited to the table.[6]

2. Power may also be used to dictate rules of the game, for instance, who gets to move first. In Figure 2 the player who moves first can dictate the outcome, provided that the other player is convinced that the first player's strategy is irrevocable.

3. Power may be used to change the payoff matrix.[7] For instance, a more powerful row player might use tactical linkage to change, or credibly threaten to change, the payoff matrix in Figure 2 to the one presented in Figure 3.[8] A large importer (read the United States) might threaten to bar imports from an exporter (read Japan) if the latter failed to make basic changes in the structure of its domestic economy, such as the distribution system.

One of the two forms of coordination (right right) is now unambiguously superior for both players, even though both forms of coordination are Nash equilibria. The players can secure the right right outcome with a

[5] Similar kinds of reasoning about commitment have been used in the recent literature on strategic trade theory. One element of this analysis is that state intervention to promote a particular industry is desirable because it demonstrates a level of national commitment that could not be provided by the action of private firms. See, for instance, James A. Brander, "Rationales for Strategic Trade and Industrial Policy," in Paul R. Krugman, ed., *Strategic Trade Policy and the New International Economics* (Cambridge: MIT Press, 1986), 30.

[6] Snidal (fn. 1), 938, points out that the threat of exclusion can itself be an effective bargaining tool. By threatening to exclude, a more powerful state might secure the compliance of a weaker state.

[7] Underlying power capabilities or size may determine the payoff matrix in the first place. See John A. C. Conybeare, *Trade Wars: The Theory and Practice of International Commercial Rivalry* (New York: Columbia University Press, 1987), 2.

[8] Albert Hirschman has argued that the credibility of such a threat would be determined by the relative opportunity costs of change; see Hirschman, *National Power and the Structure of Foreign Trade*, rev. ed. (Berkeley: University of California Press, 1980). Tactical linkage is discussed in Ernst B. Haas, "Why Collaborate? Issue-Linkage and International Regimes," *World Politics* 32 (April 1980), 357–405.

	Mountains Left	Ocean Right
Mountains Left	2, 1	0, 0
Ocean Right	0, 0	3, 2

FIGURE 3
BATTLE OF THE SEXES WITH
TACTICAL LINKAGE

ıinimum level of coordination, for instance, by playing the game in an xtended form in which either player has the option of going first. Once his outcome is reached, it is both Nash and Pareto optimal; there is no ncentive for either actor to change.

In contrast to situations involving distributional choices along the Pa- eto frontier, such as the Battle of the Sexes payoff structure, problems f collaboration, which are generated by dilemmas of common interests, re concerned with market failure.[9] They are characterized by Pareto uboptimal outcomes; at least one actor can gain without compromising he utility of others. The locus classicus of this set of problems is Prison- r's Dilemma.

A great deal of the cooperation and regime literature in international elations has been devoted to analyzing how actors can escape from a tuation in which there are incentives to cheat but in which mutual ooperation is better than mutual defection.[10] The fundamental question as been: How can players move toward the Pareto optimal outcome of tual cooperation?[11]

everal answers have emerged. Cooperation is more likely when there eration and no defined number of plays, and when discount rates are

The terms *coordination* and *collaboration* are taken from Stein (fn. 1).

Robert Axelrod and Robert O. Keohane point to Pareto suboptimality in general as the ning problem for the cooperation literature, arguing that "what is important for our poses is not to focus exclusively on Prisoner's Dilemma *per se*, but to emphasize the fun- ıental problem that it (along with Stag Hunt and Chicken) illustrates. In these games, ›pic pursuit of self-interest can be disastrous. Yet both sides can potentially benefit from eration—if they can only achieve it"; Axelrod and Keohane, "Achieving Cooperation er Anarchy: Strategies and Institutions," *World Politics* 38 (October 1985), 226–54, at 231.

Charles Lipson, for instance, argues that "because the Prisoner's Dilemma highlights the potential gains from cooperation and the temptations that prevent it, it has been n as an elegant expression of the most profound *political* dilemmas, including that of the al contract. Indeed, Jon Elster once defined politics as 'the study of ways of transcending Prisoner's Dilemma' "; Lipson, "International Cooperation in Economic and Security airs," *World Politics* 37 (October 1984), 1–23, at 3.

low and the difference between the payoffs for cooperation and defecti are modest.[12] Most important for the international relations research pı gram growing out of the literature on market failure is the investigati of the way in which cooperation can be facilitated by institutions tl reduce the temptation to lie, cheat, and dissimulate by increasing t symmetry and amount of information, raising the cost of illegitim: behavior, promoting convergent expectations, and fostering cross-issı area linkages.[13] The central normative concern of cooperation theory to develop institutional mechanisms that inhibit cheating.[14]

Market failure analyses, which have dominated the literature on i ternational regimes, pay little attention to power. Once the game is d fined, all actors are treated symmetrically with regard to their capabi ties. Both the row and the column player have the same competence play the game. No one is eliminated through brute force as opposed choosing a bad strategy. In this game, clubs are never trump.[15] Everyo could be better off if only market failure problems were solved. Power- which can be understood as the ability to determine who plays the gam or to define the rules, or to change the values within the payoff matrix- does not matter as much as information and monitoring capabiliti When analysts working within the cooperation research program ha discussed the larger question of the context within which the game specified, they have pointed to the ways in which issue linkage or limi ing the number of players might facilitate cooperation rather than tl distributional consequences of power.[16]

Global Communications

For international communications, however, market failure problem have been irrelevant. Monitoring and information were never a centr consideration. Power, not just interest, did matter, though. And th

[12] For an exceptionally intelligent and nuanced discussion of the impact of these factors the prospects for cooperation in Prisoner's Dilemma and other games, see Kenneth A. Oy "Explaining Cooperation under Anarchy: Hypotheses and Strategies," *World Politics* 38 (O tober 1985), 1–24.

[13] The most important exposition of the functions of institutions can be found in Keohan (fn. 3), chap. 6.

[14] The central importance that neoliberal institutionalism accords to the problem of che ing is elegantly elaborated in Joseph M. Grieco, *Cooperation among Nations: Europe, Ameri and Non-Tariff Barriers to Trade* (Ithaca, N.Y.: Cornell University Press, 1990), esp. chaps. 2.

[15] Grieco (fn. 14), 38, points out that one of Axelrod's premises is that there is no way eliminate a player or to avoid interactions.

[16] Axelrod and Keohane (fn. 10), 253, summarize: "We have seen that governments ha often tried to transform the structure within which they operate so as to make it possible the countries involved to work together more productively."

estion was not how to move to the Pareto frontier, but rather which int along the frontier would be chosen.

In the four issue-areas considered in this paper—radio and television ɔadcasting, remote sensing, allocation of the electromagnetic spectrum, d telecommunications (telephone and telegraph links including com- ɜnications satellites)—there are no examples of harmony. And in two ɪes, radio broadcasting and remote sensing, there are no international gimes: there has been no agreement on principles and norms because ɔre powerful states have been able to secure their first best outcome ·ough unilateral action. The critical weakness of some states in these ue-areas was their inability to regulate access to their own territory ːause they could not effectively block all broadcasting signals or re- ɔte sensing probes.

In the other two issue-areas—allocation of the electromagnetic spec- ɪm and telecommunications—there have been international regimes. th initially posed pure coordination problems; actors wanted to avoid : mutually undesirable outcomes of radio interference and incompati- · national communications systems. In recent years, however, distri- tional issues have become more consequential. Third World statesmen ve worried that the entire electromagnetic spectrum would be allo- ed without taking account of the future needs of their countries. The gime has responded to these concerns because in this case Third World ɪntries have power conferred by their ability to interfere with other tes' broadcasts and by their membership in the ITU.[17]

In the area of telecommunications, technological change has altered : capabilities of actors and increased distributional conflicts. Interna- nal regimes, in turn, have changed in response to these changes in ɔabilities. More precisely, technological innovation gave some private ors, primarily domiciled in the United States, an incentive to press for nore competitive telecommunications regime both domestically and ːrnationally. The direct and indirect economic bargaining power con- red by these new technologies gave the United States the leverage to ure some changes in the extant regime (which had legitimated na- ɪal monopolies) either through direct pressure or by changing the in- national market incentives confronting other actors.[18]

Modifications of the regime have been initiated by public threats of ection or unilateral action. Cooperation theory—which has focused

[17] This same analysis of the allocation of radio frequencies is offered by Stein (fn. 1), 131–

[18] Here and later in the paper I draw heavily on the superlative analysis of Peter Cowhey; Cowhey, "The International Telecommunications Regime: The Political Roots of Inter- ɪonal Regimes for High Technology," *International Organization* 44 (Spring 1990), 169–99.

Efforts by receiving countries to block international radio transmi sions have had only limited success. Jamming has been used almost sinc the inception of international broadcasting. Austria, Germany, and Ital jammed foreign broadcasts during and before the Second World Wa The Soviet Union has jammed since the 1930s. Spain under Franc some Middle Eastern countries, and China have also engaged in jar ming.[23]

States have also attempted to control what their citizens can hear b limiting the capacity of radio receivers. In 1951, according to one stud only 18 percent of radio receivers in the Soviet Union were capable c direct reception; the rest could receive only over wires. South Africa de veloped an ultrahigh frequency network and produced sets with onl one wave band, making it impossible to receive international broadcast Ghana at one time sold sets capable of receiving only a few stations.[24]

Efforts to impede the reception of radio transmissions have never bee fully effective however. Jamming requires both long- and short-rang stations, and, even with extensive capital investment, it is difficult t block all signals. Western broadcasts to the Soviet Union, for instanc have operated on as many as sixty frequencies and can cozy up to Sovie stations.[25] Furthermore, efforts at state control may only increase the cu riosity of potential listeners. Radio receivers are too widely available an too easily altered to permit effective state regulation.

States have even had difficulty coping with pirate stations. Northern Europe has been the most fertile ground for pirate radio stations, becaus government-based monopolies limit options for listeners. In Italy an France land-based pirates were so successful that the respective govern ments were forced to legitimate them. Spain refused to sign a European agreement that would have prohibited Spanish subjects from supplying and repairing pirate stations at sea. British regulations impeded, but did not eliminate, pirate stations in the London area.[26]

In sum, the technology of radio broadcasting has made state control of transborder radio transmissions extremely difficult. Hertzian waves do not recognize political boundaries. The technology of radio, unlike that of television, developed one universal standard: a receiver that worked in one country would also work in another. Radio signals could

[23] Ibid., 100–102; Murty (fn. 20), 56; Martin (fn. 20), 85–86.

[24] Bumpus and Skelt (fn. 22), 102; Murty (fn. 20), 55.

[25] Jonathan Eyal, "Recent Developments in the Jamming of Western Radio Station Broadcasting to the USSR and Eastern Europe," *Radio Liberty Research* RL 419/86, Radio Free Europe, Radio Liberty (1986), 2.

[26] Douglas Boyd, "Pirate Radio in Britain: A Programming Alternative," *Journal of Communication* 36 (Spring 1986), 83–94, at 86–92.

be transmitted over very long distances. Because states have disagreed about the basic principles and norms that should govern radio broadcasts and because the available technology gives senders more power than receivers, there is no international regime for radio broadcasting. Those states that favor the open dissemination of information (or propaganda, depending on one's perspective) have prevailed.

TELEVISION BROADCASTING

The international spillover of signals has been less problematic for television than for radio. This difference, however, is more a reflection of the relative power imparted by technology than of any agreement on principles and norms: television is less geographically promiscuous. As ground-based television broadcasts do not travel long distances, spillovers have been limited to contiguous geographic areas. Even with the larger footprint of direct broadcasting satellites, signals are still regionally contained.

Although terrestrial TV signals can travel only a few hundred miles, there have been at least twenty-four cases in which broadcasting affected another country. The United States licensed TV stations aiming at a primarily Canadian audience. Israel initially tried to keep out television entirely because it was regarded as lowering cultural standards, but the government ultimately capitulated because Israeli citizens were receiving broadcasts from stations in Arab countries that carried many American programs. Hungarian and Czechoslovakian televisions were converted so that they could receive broadcasts from neighboring Western countries, a more complicated procedure than altering a radio, but one that could be performed by many television technicians. Most significantly, West German television could be received in most of East Germany. It was so popular that the East German government was compelled to install cable around Dresden, the one part of the country that could not directly receive Western broadcasts—this, to attract workers to the area.[27]

The development of direct-broadcasting satellites (DBS) posed more bstantial challenges to national control. Initially, DBS, like radio, ap-

[27] George Quester, *The International Politics of Television* (Lexington, Mass.: Lexington oks, 1990), 30, 105, 113–14, 126–30. One poll of East German refugees (not exactly a ranm sample of the East German population) taken in 1985 indicated that 82 percent watched est German television on a regular basis. See Dieter Buhl, "Window to the West: How levision from the Federal Republic Influenced Events in East Germany," Discussion Par D-5 (Cambridge: Joan Shorenstein Barone Center, Kennedy School, Harvard University, 90), 3. Buhl argues that West German television undermined the legitimacy of the comnist regime.

peared to be a technology that could circumvent state control, becaus any individual with a television receiver could access transmissions fron other countries. This fear has proven to be unfounded, though; interstat agreements have effectively regulated most satellite television broadcast:

International agreements have endorsed the principle that televisio signals should be limited as much as possible to national markets. Th 1971 World Administrative Radio Conference (WARC) on space telecom munications concluded in part that DBS spillover should be minimize unless there was an explicit agreement among countries. WARC dele gates have agreed to utilize different channels, different orbital position and different polarizations for their transmissions from direct-broad casting satellites. If these provisions were breached there would be inter ference with the broadcasts. Implied by the WARC agreements, therefore was the de facto acceptance of the principle of no broadcasting withou prior consent.[28]

Agreement, however, has been limited to areas where states hav shared interests and relatively equal power. Europe offers the prime ex ample. European states have authorized the expansion of direct-broad casting satellites as one manifestation of a more integrated Europe; the have also, however, imposed restrictions on advertising and content an encouraged the development of more European programming. Com menting on American objections to a Community directive that at leas 50 percent of programming—excluding sports, game shows, and adver tisements—be produced in Europe, Jacques Delors, the president of th European Commission stated: "I say to the United States, 'Have we th right to exist, to perpetuate our traditions?' "[29]

In contrast, there has been no agreement on broadcasting between Cuba and the United States. The United States has beamed television signals at Cuba, and the Cuban government has jammed them. Like wise, West Germany consciously acted to penetrate East Germany. When East Germany adopted a color TV system that was incompatible with West German broadcasts (the black and white systems had been th same), the West Germans began broadcasting on both their own and th East German standard.[30]

In sum, land-based television broadcasting has presented less of a

[28] Stephen Gorove, "International Direct Television Broadcasting by Satellite: 'Prior Con sent' Revisited," *Columbia Journal of Transnational Law* 24 (1985), 8, 58. See also David E. S Blatherwick, *The International Politics of Telecommunications*, Research Series (Berkeley: In stitute for International Studies, 1987), 42, 47.

[29] *Europe*, April 1989, p. 29; *Financial Times*, June 15, 1989; *New York Times*, October 4 1989, p. C17.

[30] Quester (fn. 27), 126.

problem than radio because the available technology has limited reception; nevertheless politically controversial spillovers did occur in the Middle East and Europe. When direct-broadcasting satellites increased the footprint of signals, states with the same preferences and the mutual ability to interfere with each other's broadcasts, such as the members of the European Community, could coordinate, but those with disagreements, such as Cuba and the United States, could not. Where disagreements on the principle of prior consent existed, the payoff matrix was Deadlock. There was no agreement because for both parties the first best choice was unilateral uncoordinated action. Those states favoring open access prevailed because the distribution of power, derived from the available technology, made it easier to transmit successfully than to block radio and television transmissions. The issue was where states would end up on the Pareto frontier, not how to get there.[31]

Remote Sensing

Remote sensing is another issue-area in which there is no agreement on basic principles and therefore no foundation for an international regime. Remote sensing involves electronic information gathering from satellites. States have disagreed about basic principles and norms—the right to sense versus the need to secure prior consent from the target state. Power is asymmetrically distributed: target states have not been able to block sensing probes. The behavior of those more powerful states that are able to engage in remote sensing has been constrained only by the need to secure territorial access to make use (usually commercial) of some of the data secured from satellite probes.

The first American remote sensing satellite was launched in 1972. A number of countries, including India, China, France, and Japan, have their own programs. These satellites can gather important strategic and commercial data from anywhere. It is very difficult, and in many cases impossible, for sensed states to disrupt this process (as distinct, say, from the possibilities in systems with gateways such as oceanic cables that are easily accessible to state authorities). While some commercial applications of data gathered by remote sensing, such as the exploitation of minerals, do ultimately require territorial access, other kinds of information, such as the deployment of military forces or the prospects for crop yields, do not.

The international rules that have evolved in the area of remote sensing

[31] Assuming no agreement on principles, the future of state control will depend on power derived from technological choices. It would, for instance, be easier to regulate transmissions sent through fiber-optic cables than those sent via satellites. Ibid., 138–39

are *not constraining*.[32] The UN General Assembly has endorsed the fol lowing principles: remote sensing is not to be carried out in a way tha would be detrimental to the sensed state; sensed states are to be allowe to participate in programs and are to be given access to processes as wel as raw data; sensing states are to notify the UN of their programs. Th United States, the Soviet Union, and Japan, however, have stated tha they do not regard these rules as legally binding.[33]

A number of possible regimes were rejected. Developing countrie had initially endorsed the view that there should be no remote sensin without the prior approval of the sensed state.[34] But without the abilit to block satellite probes, states could not enforce such a rule or compe others to accept it. Sweden had advocated the creation of an internationa regime similar to that for satellite communications, in which control ove the gathering and dissemination of remote sensing data would be subjec to a formal international agreement. This too failed.[35] Unlike telecommunications, where the technology would be useless or its benefits severely curtailed without state-authorized territorial access, remote sensing allowed actors in one state to secure useful information abou conditions in another without formal approval.

Nonetheless, remote sensing has not entirely evaded the control of sensed states. From the outset, the United States accepted the fact that the commercial benefits of its remote sensing program would be limited unless it engaged the support of other countries. It backed the development of local receiving stations and made the raw data secured from its LANDSAT available to all comers at affordable prices, an approach that coincided with the traditional American commitment to the free flow of information. Not surprisingly, this openness was not extended to the security arena; satellite photographs with very fine resolution remain highly classified.[36]

Hence in the area of remote sensing, as with radio and television broadcasting, there is no agreement with regard to the basic principle of prior consent, that is, the right of a national state to control access to its own territory. Remote sensing is an area in which more powerful states have not confronted problems of coordination or distribution. Consequently, their first best solution was to act unilaterally; they have provided target states with some information because the latter had the power to control the territorial access necessary to realize some of the commercial benefits of remote sensing.

[32] Marvin Soroos, *Beyond Sovereignty: The Challenge of Global Policy* (Columbia: University of South Carolina Press, 1986), 340.
[33] Blatherwick (fn. 28), 72–76.
[34] Ibid., 62–64. [35] Ibid., 74–77. [36] Ibid., 57, 77.

'HE ELECTROMAGNETIC SPECTRUM

ı the last two issue-areas related to international communication con-dered in this paper—allocation of the electromagnetic spectrum and :lecommunications—no state could achieve its objectives through urely unilateral action. The allocation of the electromagnetic spectrum ıvolves a classic coordination problem. If there is no general agreement n frequency allocation, then broadcasts can interfere with each other. 'his can happen even if states agree on the principle of the open dissem-ıation of information without prior consent. Distributional issues, 'hich are also a factor, include determining which groups of roadcasters should have standing, allocating geosynchronous orbit slots ›r broadcast satellites, and, most important, defining the principle that ıould be used for partitioning the electromagnetic spectrum. Since the ›ectrum is a limited resource, who gets how much and why?

The International Telecommunication Union (ITU)—an intergovern-ıental organization and the various World Administrative Radio Con-·rences (WARCs) that it sponsors—has been the venue within which these sues have been resolved. States have played the decisive role because ıey are the only actors capable of enforcing agreements, that is, of con-'olling broadcasts emanating from within their own territory.

The major question associated with the electromagnetic spectrum has een the determination of principles and rules upon which allocation ıould be based. Occasionally this matter has been resolved on technical rounds; for instance, the spark wireless sets used on ships at the begin-ing of the twentieth century were inexpensive but very wasteful of the ıdio spectrum. They were eventually phased out.

Most competing claims have not been so easily settled, however, for ıstance, the allocation of geosynchronous orbits, which are the most ef-cient locations for broadcasting satellites. The claims of some equatorial ates to sovereign control of slots above their own territory have simply :en ignored by more powerful states. The allocation of these slots has :en made at World Administrative Radio Conferences dealing with ›ace orbits. These deliberations have given INTELSAT, an international ›nsortium that controls most communications satellites, priority rights ver national and binational satellite systems. INTELSAT had the sup-ɔrt of Third World and other countries.[37] The regime clearly mattered :re because the venue for making this decision, the ITU, which sponsors 'ARCs, is a universal international organization based on the principle of ›vereign equality: one nation, one vote. The issue here is not whether

[37] Andrea Kavanaugh, "Star WARCs and the New System: An Analysis of U.S. Interna-ınal Satellite Policy Formation," *Telecommunications Policy* (June 1986), 93–106, at 105.

the Pareto optimal frontier would be reached but which point along the frontier would be chosen.

In dividing up the spectrum there have been two rival positions. The first, which dominated allocation until very recently, has been that frequencies should be designated on a first-come-first-serve basis, provided claimants could make effective use of the resource. The second, championed by less developed countries, has been that at least some portion of the spectrum should be allocated on the basis of sovereign equality. The concern of the developing countries is that by the time their needs increase, already developed countries will have claimed virtually all usable frequencies.

The principle of basing allocation on usage was agreed to very early. The 1906 convention on radiotelegraphy specified that certain radio bands be used for certain types of services. Radio frequencies were registered on a first-come-first-serve basis. Countries notified the Bureau of the International Telegraph Union (the predecessor of the International Telecommunications Union) in Bern that they had established a communications channel. Other countries, in turn, were prohibited from interfering with stations that were on the Bern list.[38] By 1980 this first-come-first-serve set of rules had resulted in a situation in which the Soviet Union and the United States claimed half of the available frequencies and 90 percent of the spectrum was allocated to provide benefits for 10 percent of the world's population.[39]

Developing countries have challenged rules of the game based solely on present need, arguing that some part of the spectrum should be reserved for future demand. Delegates to recent WARCs have accepted this principle. Developing countries derived leverage from the fact that, as states, they had a presumptive right to participate in ITU conferences, which afforded them access to the relevant decision-making forums and facilitated coordination of their strategies. Unlike many other areas, in which LDCs have little or no directly relevant power capabilities (they cannot with two or three exceptions, for instance, launch satellites, nor can they shoot them down), developing states can interfere with the signals of neighboring countries.

The political issues involving the international regime for allocating the radio spectrum were not generated by market failures leading to Pareto suboptimal outcomes (the problem that has informed most theo-

[38] George A. Codding, Jr., and Anthony M. Rutkowski, *The International Telecommunication Union in a Changing World* (Dedham, Mass.: Artech House, 1982), 13, 25–26.

[39] Stephen D. Krasner, *Structural Conflict: The Third World against Global Liberalism* (Berkeley: University of California Press, 1985), 229.

ies about international regimes). Rather, they grew out of matters in-
olving coordination with distributional consequences. Unilateral action
vould leave all actors worse off, but any form of coordination would
rivilege some over others. This is a classic Battle of the Sexes payoff
natrix. The point along the Pareto frontier that has been chosen—one
t which most frequencies are still allocated on a first-come-first-serve
asis that benefits industrialized countries—reflects the existing distri-
ution of power capabilities as manifested by the present demand for
ommunications channels. Nevertheless, the preferences of developing
ountries have not been completely ignored; unlike the cases of radio
roadcasting and remote sensing where targets had little power, in the
ase of the electromagnetic spectrum Third World states could exercise
ome leverage through their votes in the ITU and their ability to interfere
vith broadcasts from other states.

TELECOMMUNICATIONS

nternational telecommunications—telegraph, telephone, and transbor-
ler data flows—pose a fundamental coordination problem: assuring the
conomic and technical compatibility of international links. Choices also
ave distributional consequences. In recent years the most controversial
spect of these choices has concerned which actors could participate in
lobal telecommunications and whether prices should be set by market
ompetition or administrative fiat. The major impetus for change came
rom technological innovations that altered the interests, and ultimately
ne power, of the various public and private players. These technological
hanges led some actors, especially in the United States, to press for a
nore market-oriented regime, as opposed to one that legitimated na-
onal monopolies.

Despite astonishing technological changes beginning with the inven-
on of the telegraph in 1836 and continuing through the initial deploy-
nent of high-quality fiber-optic cables and satellites in the 1960s, the
nternational regime for telecommunications remained fundamentally
able. It was a regime based on national monopolies, usually public post
nd telephone and telegraph agencies (PTTs) but occasionally privately
wned firms, as with American Telephone and Telegraph (ATT) in the
Jnited States. Prices were set through administrative fiat both nationally
nd internationally. The sending country could charge whatever it
vanted for a call and paid a fixed fee to the receiving country. Senders
vere prohibited from routing calls along cheaper paths or from selling
locks of calls, which together essentially precluded price competition.
Vithin countries, at least in the case of many industrialized countries,

suppliers of telecommunications equipment were usually limited to r tional entities such as Western Electric, the equipment-manufacturi subsidiary of ATT before the latter was broken up.[40]

The history of international organizations dealing with telecommu nications began with the International Telegraph Union, which was cre ated after a meeting called by Napoleon III in 1865. (The name wa changed to the International Telecommunications Union in 1932, when radio transmission was formally added to its portfolio.) The inaugural session dealt with such coordination problems as designating hours for sending and receiving, agreeing on a common code (the international Morse code), and setting technical standards. International cable links (the first telegraph cable was laid under the English Channel in 1851) were jointly owned by the operating entities at the termini of the cables. Like cables, international wireless communication, which was first developed at the end of the nineteenth century, had to be tied into national communications networks at a limited number of gateways, to provide telegraph or telephone service. States jealously guarded their right to regulate wireless communication, even between ships. National regulations frustrated an effort by the Marconi company, the original developer of such equipment, to set its own international rules.[41]

The International Telecommunication Union (ITU) and INTELSAT, the consortium created to launch and operate communications satellites, were designed to reinforce this system. INTELSAT was a common carrier for common carriers, not a rival for established national monopolies. Peter Cowhey has argued that the "regime was in fact a political invention so successful that it eventually disappeared from sight."[42] The old regime legitimated a system of national monopolies and administered prices so successfully that the ITU and its committees came to be viewed as little more than agencies for dealing with questions of technical compatibility.

In recent years the fundamental principles and rules of the old regime—national monopolies and administered prices—have been challenged by new actors with new power capabilities, the result of techno-

[40] Cowhey (fn. 18), 177–80. See also William J. Drake, "Asymmetric Reregulation and the Transformation of the International Telecommunications Regime" (Unpublished paper, Department of Communications, University of California, San Diego, August, 1989), 3–4, 11–15. A revised version will appear in Eli Noam and Gerard Pogerel, eds., *Asymmetric Deregulation: The Dynamics of Telecommunications Policies in Europe and the United States* (Norwood, N.J.: Ablex, forthcoming).

[41] Codding and Rutkowski (fn. 38), 12–13. See also Mildred Feldman, *The Role of the United States in the International Telecommunication Union and Pre-ITU Conferences* (Baton Rouge, La.: Mildred L. Bos Feldman, 1975), 23–29, 48; and Sara Fletcher Luther, *The United States and the Direct Broadcast Satellite: The Politics of International Broadcasting in Space* (New York: Oxford University Press, 1988), 19.

[42] Cowhey (fn. 18), 169.

cal change. Telecommunications is eroding the distinction between ce and data forms of transmission. The microchip revolution created ent new equipment manufacturers. Communications systems have n linked with computers. Fax machines have obscured the difference veen conventional mail and telecommunications.

New technologies gave large users and new equipment producers an ncentive to change the extant system of national monopolies. Large isers could reduce their communications costs by developing alternatives o those services offered by national PTTs. The producers of new equipnent could realize greater economies of scale by accessing foreign as well is domestic markets.[43] More efficient telecommunications services could enhance the competitiveness of particular companies or even national economies. Unlike even the dramatic technological changes of the 1830s hrough the 1950s, the more recent developments have made it possible or individual companies to achieve greater efficiencies by developing heir own telecommunications systems rather than relying entirely on national monopolies.[44]

It is not surprising that changing incentives have changed national policies. The United States has been at the forefront of the demands for a more competitive international system, one that would be based more on market competition than on public or private monopolies imposing state-authorized tariffs. Many of the large users and manufacturers that would benefit from a more competitive environment are domiciled in he United States. Moreover, the United States had already deregulated ts domestic regime for telephone communications. Once the United States opened its own market to foreign manufacturers of equipment, it had an incentive to pry open the markets of its competitors as well.[45]

The United States has not been alone in pushing for a more competiive international environment. The United Kingdom and Japan have also supported change. Like the United States, these two countries are hubs of global communications and the home countries for major international financial and manufacturing companies that can derive large savings by introducing new communications facilities and services. The United States, the United Kingdom, and, to a lesser extent, Japan have supported what Cowhey has called the big bang—a major move away from national monopolies.[46]

[43] Ibid., 188; Drake (fn. 40), 40–42, 46–48.

[44] For instance, beginning in the 1960s IBM developed its own communications system based on satellites and lines leased from PTTs. This system now extends to 145 countries; *Business Week*, special issue no. 3033–44, January–March 1988, p. 141.

[45] Jonathan David Aronson and Peter F. Cowhey, *When Countries Talk: International Trade in Telecommunications Services* (Cambridge, Mass.: Ballinger, 1988), 218–23.

[46] Cowhey (fn. 18), 191–95.

There was initially little support for the Anglo-American position. The European PTTs were not anxious to open their national markets to competition. But some of the large European companies supported change, if only because they feared that national restrictions would give an advantage to their American competitors, who worked in a more market-oriented system. The European Community eventually supported American initiatives to place services, including telecommunications, in the GATT negotiations, where the prevailing principles and norms were market-oriented, rather than solely within the purview of the ITU, which had historically legitimated national monopolies. Even within the ITU there has been some change. A 1989 agreement recognized the right of new entities, such as IBM, to operate under the same rules as established private phone companies.[47]

In response, the PTTs and their national governments have not simply acceded to demands for a purely market-oriented regime. They have used regulations and standards to maintain their national control, for example, prohibiting the connection of private lines to publicly switched networks, banning independent transmission facilities, and requiring PTT equipment for initial connections. Because of anxiety about budget deficits, pressure from manufacturers of new equipment, and perhaps the insistence of the United States, the Japanese government has privatized Nippon Telephone and Telegraph (NTT) and permitted some competition for international services, but it has not yet endorsed a free and open market. The EC is committed to introducing a free market for terminals, private branch exchanges (PBXs), and value-added services as part of the 1992 program. PTT monopolies may be limited to such basic services as telephones.[48] Although the Canadian system is more open than those in Europe, it has more restrictions than that of the United States.[49]

Changing national policies in the area of telecommunications reflect the way in which technological innovation can alter national power and policies. Once domestic deregulation had eliminated ATT's long-distance monopoly, the company's interest in preserving monopolies in other countries disappeared. Indeed, many American producers and users would be better off under a more competitive international system, but the United States could not impose a competitive system on other countries. It could, however, pressure other actors, most obviously by threatening to block access to the very large American market for equipment

[47] Ibid., 196; Drake (fn. 40), 49, 66.

[48] Hans Bauer, "Telecommunications and the United European Market," *Telecommunications* 25 (January 1990), 33–35.

[49] Aronson and Cowhey (fn. 45), 162, 178–80.

but also by adopting national regulations that would leave more restricted foreign telecommunications users at a competitive disadvantage.

In sum, the international regime has changed. Although national monopolies still control basic domestic services in most countries, a wider range of telecommunications facilities has been introduced. There has been movement toward a more market-oriented international regime. A few large countries, especially the United States, have changed the incentives offered to other actors in the international system. Whatever the final outcome of the movement toward a more market-oriented system, the issues involved have not been ones of market failure. Nor is this a situation of harmony in which unilateral choices maximize both individual and collective interest. The tough bilateral and international negotiations over the rules governing the provision of telecommunications equipment and services suggest that while states need to coordinate their behavior, they also differ in their views about which is the most desirable point on the Pareto frontier.

Satellite Communications

The case of the development and deployment of communications satellites raised issues of coordination yet again. On the one hand, the absence of coordination would have left all actors worse off; on the other hand, the choice of which form of coordination to adopt would inevitably have distributional consequences.

The new international regime for communications satellites departed from existing practices in significant ways. International cables, the first means of transmitting messages across open waters, were usually jointly owned by various national operating entities. For satellites, INTELSAT, a new entity organized as a consortium of member states, was created. Instead of placing communications satellites under the purview of the ITU, where one-nation-one-vote practices would have given Third World and Eastern bloc countries some leverage, the new consortium distributed votes according to usage. This gave the heaviest users the largest number of votes. INTELSAT itself, rather than national operating entities, owned satellites, a departure from the practice governing international cables. Nevertheless, INTELSAT reinforced rather than challenged the prevailing system of national monopolies. Satellites that had at least the potential for point-to-point communication were not initially allowed to disturb the prevailing regime of national monopolies.[50]

The United States, the prime mover in the creation of INTELSAT, had

[50] Cowhey (fn. 18), 181.

originally hoped to handle satellite communications through a series of bilateral agreements that would have maximized the bargaining power of the United States, with its substantial technological lead in the early 1960s. The Europeans countered with a proposal to create a new international organization that would constrain American influence. The idea of a consortium, the institutional arrangement that was ultimately accepted, was also proposed by the Europeans.

Given a new set of problems in which the power resources of the actors were different from those affecting radio and cable communications, the U.S. and the major Western European states acted on their own to create a new organization rather than rely on an existing entity such as the ITU or the United Nations Committee on the Peaceful Uses of Outer Space. The consortium reflected a distribution of power in which the United States controlled vital technology but other states still regulated access to their national systems. The Soviet Union and LDCs were excluded from the opening negotiations, although the United States did conduct bilateral talks with the Soviet Union. Japan, Australia, and Canada played peripheral roles.

In June 1964 the participants concluded an interim five-year agreement. Ownership and voting power would be based on usage, which gave the United States 61 percent and Western Europe 30.5 percent of the votes. Australia, Canada, and Japan received 8.5 percent. Quotas could be adjusted to give developing countries up to 17 percent, but no more. These arrangements allowed the United States to maintain a secure controlling interest—but with limitations. If there was disagreement on important issues, voting rules mandated that passage required 12.5 percent of votes over and above the votes of the member with the largest number of votes. Hence, the United States could not act alone.[51]

Nineteen official actors (eighteen countries and the Vatican) signed the final agreement in July 1964. Membership in INTELSAT has since grown to more than one hundred countries. Voting power is still based on usage and is reallocated once a year on the basis of activity over the previous six months.[52]

The national power capabilities provided by technological and financial resources initially gave the United States the dominant role in INTELSAT, even if American policymakers failed to secure their first best insti-

[51] Joseph N. Pelton, *Global Communications Satellite Policy: Intelsat, Politics and Functionalism* (Mt. Airy, Md.: Lomond Books, 1974), 54–59.

[52] Richard R. Colino, "Global Politics and INTELSAT: The Conduct of Foreign Relations in an Electronically Wired World," *Telecommunications Policy* 10 (September 1986), 199.

tutional preference of bilateral arrangements. Not only did the United States have more voting power, but its designated operating entity, Comsat (the Communications Satellite Corporation established as a joint venture for American participation in communications satellites), was recognized in the 1964 Interim Agreement as the operational manager for INTELSAT. Daily operations were entirely in Comsat's hands. No INTELSAT secretariat or international civil service was specified by the Interim Agreement.[53] Over time, the influence of the United States eroded, as the underlying power capabilities of other states increased with the diffusion of technology. Negotiations for a permanent agreement to supplant the 1964 Interim Agreement took place between 1969 and 1973. These discussions resulted in a decrease in American control including a reduction in the role of Comsat.[54]

In more recent years INTELSAT has actually hindered American efforts to create a more competitive market-oriented regime by developing alternatives to INTELSAT's satellites. The director general of INTELSAT argued in a 1986 article that

> one other conclusion which may be reached is that foreign policies and relations should not be the product of a "Johnny-one-note" economic philosophy. U.S. estrangement from INTELSAT is not a desirable or viable alternative for world peace and understanding. And, while the currently popular economic trilogy—deregulation, pro-competition, and private sector predominance—may have its value and applicability, it is simply insufficient to serve as a complete litmus test for enlightened and effective foreign relations.[55]

The United States itself has approved some new satellite operations, but with substantial restrictions that recognize the central role of INTELSAT. The FCC ruled that new operations could not connect with public switching networks, because they would have directly competed with INTELSAT's services. New systems could be established only with the permission of a foreign government and if they went through an INTELSAT review of coordination.[56]

Alternative satellite systems would be useless so long as states have effective control over gateways to domestic networks. As in the case of cables, the acceptance of INTELSAT competitors will depend upon national calculations of interests. Central decision makers in the United States have endorsed private initiatives. State actors in other technologically advanced countries may do the same. International competition will give

[53] Pelton (fn. 51), 82.
[54] Ibid., 123; Kavanaugh (fn. 37), 97.
[55] Colino (fn. 52), 208.
[56] Aronson and Cowhey (fn. 45), 121–32.

states with more limited technological capacities an incentive to accept non-INTELSAT satellite systems lest their economies be placed at a competitive disadvantage. With the erosion of INTELSAT's monopoly, the international regime for communications satellites will become more appealing for some states and less so for others.

The development of communications satellites created a new set of technological challenges and new problems of distribution and coordination among states and operating entities. The resolution of these problems was consistent with the distribution of power in this issue-area. Without active American participation, the development of communications satellites would have been impossible in the first place; the United States initially enjoyed a dominant position because of its command of technology. Other states preferred a satellite system in which the United States wielded the greatest influence to no system at all, but they were still in a position to force some compromises because they still controlled access to their own national communications networks, and without such access satellites would be useless. Communications satellites were regulated by a consortium not a set of bilateral treaties. Moreover, as American technological capabilities declined relative to those of other industrialized states, U.S. influence in INTELSAT eroded. As in the area of telecommunications more generally, the United States has had some, but far from total, success in pushing for a more market-oriented regime. The main incentive for acceding to American pressure has been the fear of being placed at a competitive disadvantage. The main instrument for resisting such pressure is the ability of states to regulate access to their own systems.

The issue in the case of communications satellites was not market failure or monitoring adherence to agreements. Rather, it concerned the distribution of rewards. There are many points along the Pareto frontier. Those that have been chosen have reflected the relative bargaining power of states.

Conclusions

The literature on international regimes has focused on market failures. It has emphasized the possibility of mutual gains. When the logic of individual utility maximization leads to Pareto suboptimal outcomes, appropriate institutional constructs can enhance well-being. Power, the conventional focus of the literature on international relations, is not necessarily inconsistent with this perspective. Robert Keohane has argued that what he calls the contractual approach is not a substitute for a

power-oriented or interdependence position, but is rather a supplement to these "traditional modes of political analysis."[57]

Research programs, however, have both a denotation and a connotation. While the denotation, or explicit logic, of a research program based on the investigation of market failure is not inconsistent with a power-oriented analysis, the connotation of this research program is that power can be ignored. The connotation of a research program suggests which questions are most important, what kind of evidence should be gathered, and, often tacitly, which issues should be ignored. In a penetrating analysis of the positive theory of institutions, Terry Moe argues that

> over the last decade, as the positive theory of political institutions has developed as an intellectual enterprise ... [t]he theme that gets driven home again and again is that political institutions help mitigate collective action problems—and that this is why they exist and take the forms they do. While the positive theorists are well aware that politics is a game of winners and losers and that institutions are crucial means by which winners win and losers lose, this is not an equally important theme. All too often, it is not discussed at all.[58]

The research program that emerges from an emphasis on market failure, one specific manifestation of the problem of social choice, suggests that the most important issue is how to reach the Pareto frontier. Prisoner's Dilemma is the exemplary payoff matrix, but other situations such as Stag Hunt or Chicken, all of which can result in Pareto suboptimal outcomes, are also examined. Deadlock, the one other game that has been mentioned with any frequency in the cooperation under anarchy literature, is not extensively discussed because it is not analytically interesting. Battle of the Sexes is hardly noted at all as a possible payoff matrix.[59] Analysis involves investigating the kinds of institutional arrangements that can overcome the problem presented by a payoff structure in hich the Pareto optimal outcome is not a Nash equilibrium. For exnple, although Keohane explicitly argues in *After Hegemony* that both ›wer and exchange determine outcome, the most heuristically compelıg section of the book, chapter 6, examines the ways in which institu›ns can overcome problems of market failure by, for instance, increasg the symmetry and amount of information, making it easier to tablish reputation across several issue-areas, and promoting convergent

[57] Keohane, "Multilateralism: An Agenda for Research," *International Journal* (Autumn ›o), 746. See also Keohane (fn. 3), 21.

[58] Moe, "Political Institutions: The Neglected Side of the Story" (Paper prepared for the le Law School *Journal of Law, Economics, and Organization* Conference on the Organizan of Political Institutions, April 27–28, 1990), 1–2.

[59] For an excellent discussion of various payoff matrices, see Oye (fn. 12), 12–18.

expectations. It is this chapter that has influenced the research program on cooperation, not Keohane's recognition of the fact that institutions are created within a power-driven context.

Power is ignored because the research program based on the analysis of market failure does not provide either heuristic insights or analytic techniques for investigating relative capabilities. Power recedes into the background not only because scholars have studied absolute rather than relative gains,[60] but also because it is not evident how power is relevant for solving problems of market failure. If the purpose of international regimes is to enhance both the amount and the symmetry of information so that states can be more confident that they can prevent cheating, then the cognitive ability to construct efficacious institutional arrangements is more important than national power capabilities. Cleverness can make everyone better off.

Conversely, a power-oriented research program is not logically inconsistent with the analysis of market failure. But again, the connotation of such a research program pushes issues of Pareto suboptimality, cheating, and monitoring into the background. The most important issue for a power-oriented analysis is the distribution of capabilities and benefits. Charles Perrow, for instance, argues that there is always a struggle within an institution because control of the institution can bring with it a variety of rewards including security, power, and survival.[61] For a power-oriented research program, power is exercised not to facilitate cooperation but to secure a more favorable distribution of benefits.[62] And analysis seeks to explain outcomes in terms of interests and relative capabilities rather than in terms of institutions designed to promote Pareto optimality.

The primary motivation for establishing international regimes for international telecommunications has been to solve coordination problems that have distributional consequences—not to address problems of market failure. States wanted some set of rules for the allocation of the electromagnetic spectrum and international telecommunications including satellites, because the failure to coordinate policies on noninterference and on the compatibility of national networks would have left everyone

[60] This point is forcefully made in Grieco (fn. 2).

[61] Perrow, *Complex Organizations: A Critical Essay*, 3d ed. (New York: Random House, 1986), 132. Perrow criticizes economic analyses, such as the principal agent literature, for ignoring the distribution of power (pp. 230, 257–58).

[62] Contrast (1) the discussion in Axelrod and Keohane (fn. 10), 249, about the way in which background conditions, such as issue linkage, can be altered to facilitate cooperation with (2) the analysis in this paper, in which power is used to alter background conditions (players, issue linkage, payoff matrix) to change the distribution of benefits.

vorse off. But because the decisions would be consequential for the dis-ribution of rewards, conflict arose over how such coordination problems hould be resolved, that is, where along the Pareto frontier they would nd up. While all actors were better off with some form of coordination ather than none, the form of coordination adopted would affect them lifferently.

The regimes that have been established for the allocation of the radio pectrum and international telecommunications have reflected the rela-ive power of states and have changed as the distribution of power has hanged. Regimes were not irrelevant; without at least some shared rules ll actors would have been worse off. Power has been determined by hree considerations: technology and market size, which have influenced he relative opportunity costs of change and therefore the ability to make redible threats; membership in universal international organizations, vhich has given states the presumptive right to influence policies that re affected by one-nation-one-vote decision-making procedures;[63] and ontrol over territorial access provided by juridical sovereignty.

For instance, INTELSAT became much less of an American preserve as he technological capabilities of other states improved. Third World tates were able to secure some changes in the first-come-first-served rinciple of allocating the electromagnetic spectrum because they had a resumptive right to membership in the ITU and because, once they had heir own broadcasting facilities, they could interfere with the radio sig-ials of others. The United States was able to secure some movement oward a more market-oriented regime for global telecommunications ecause some actors in other states feared that they would be put at a ompetitive disadvantage by unilateral American action, but U.S. lever-ge was limited by the fact that state authorities controlled access to their wn national networks.[64]

In two cases—radio broadcasting and remote sensing—there has been o agreement with regard to the basic issue of whether sending states iave to secure the prior consent of targets. States opposed to prior con-ent could secure a better distribution of rewards through unilateral ac-ion, that is, by broadcasting or sensing without the prior consent of tar-et countries rather than by attempting to coordinate with those opposed o the free dissemination of information. Third World states secured

[63] This argument is elaborated in Krasner (fn. 39).

[64] Such control is not foreordained even for telephone communications. While interna-onal links for both satellite and cable are now connected to national systems through a mited number of easily regulated gateways, Motorola has announced plans to develop a ortable telephone system that could send and receive calls point to point anywhere on earth y bouncing signals off seventy-seven satellites; *New York Times*, June 26, 1990, p. 1.

some data from remote sensing because their approval for te
cess was needed to realize the commercial benefits of some in

There has been no systematic classification of internation
with regard to whether market failure, coordination, or dis
questions have been central. Are interactions among states b
terized by Prisoner's Dilemma, Stag Hunt, Deadlock, the B
Sexes, or some other payoff matrix? The literature on coop
focused on payoff matrices in which cheating is a central p
cause the Pareto optimal outcome is not a Nash equilibriun
search program is academically attractive because it allows p
entists to deploy a heuristically powerful array of analytic tools
by economists, not to speak of the appeal of identifying with
higher on the social science pecking order and better able to p
as a true science, Nobel prize and all.[65]

How important is market failure for the study of internat
tions? It is illusory to suppose that this issue will be resolved o
of empirical studies. The literature now contains enough e
suggest that both market failure and power-oriented research
can present plausible analyses, often about the same issue.

There are, however, three aspects of the international syster
gest that a power-oriented approach is, in many cases, more a
than analyses based on market failure. First, there are som
international politics, especially but not exclusively related t
that are zero sum. What is at stake is the power, that is, t
capacity, of actors. Power-oriented concerns may be directed
ward altering the behavior of others or toward preserving
autonomy. Market failure is never at issue here; one actor's
other's loss.[66]

Second, in international relations it is possible to eliminate
ers through the use of force or to compel an actor to accept a
that it would never agree to voluntarily.[67] This is an option
analytically tractable for a market failure research program,
sumes that actors are in a position to make voluntary ch
Austro-Hungarian Empire did not chose to dismantle itsel
First World War. Munich compelled Czechoslovakia to acce
come that left it absolutely worse off.

Third, even if states are interested in absolute rather th

[65] Robert Jervis has also argued that one of the attractions of Prisoner's Dilen
lends itself to interesting manipulations; see Jervis, "Realism, Game Theory, a
tion," *World Politics* 40 (April 1988), 317–49, at 323.

[66] See ibid., 334, for a similar formulation. Grieco (fn. 14) places particular
the desire of states to preserve their freedom of action.

[67] I am indebted to Terry Moe for pointing out this line of argument.

gains, the initial allocation of property rights will have distributional consequences, even if any particular specification of such rights can lead to a Pareto optimal outcome.[68] Different choices will differentially privilege different actors.[69] Changing the rules for using the radio spectrum, or setting tariffs for INTELSAT, or lending IMF resources, or assigning landing rights for international airlines, or limiting whale catches would benefit some actors at the expense of others. The issue here is movement along the Pareto frontier, not how to reach the frontier. Power and interests, not monitoring capacity, determine outcomes.[70] These three characteristics—concern with relative power, the ability to eliminate actors, and the assignment of property rights—are the core of much of the literature on international relations.

This is not to say that institutions or market failure problems are irrelevant. First, as John Ruggie, in particular, has elegantly argued, the existence of the basic actors in the contemporary international system—national states—must be understood in a larger, historically grounded institutional context. The issue here, however, is not the temptation to cheat and the ability to monitor but rather an understanding of how the players in the game came to exist in the first place.[71] Second, the free-rider problem that emerges when there is a very asymmetrical distribution of power will lead to the Pareto suboptimal provision of collective goods. In these circumstances, however, the problem is not cheating and monitoring but rather the fact that the more powerful state (or small group of states) prefers to provide the collective good regardless of what the smaller states do, and the smaller states prefer to defect regardless of what the larger state does. A situation of Prisoner's Dilemma arises only if there are a larger number of smaller states, no one of which is willing to provide collective goods unilaterally. Third, past institutional choices do influence the contemporary interests and power of actors exactly because such choices privilege some players at the expense of others.

[68] Ronald Coase, "The Problem of Social Cost," *Journal of Law and Economics* 3 (1960), 1–4. John Conybeare has applied Coase's logic to problems of international organization. See Conybeare, "International Organization and the Theory of Property Rights," *International Organization* 34 (Summer 1980), 307–34, esp. 322–23, where the author recognizes, although does not elaborate, the income effects of the initial distribution of property rights.

[69] For instance, Joseph Grieco's recent study of the nontariff barrier codes negotiated during the Tokyo Round suggests that conventional concerns about relative power are a more persuasive explanation of outcomes than neoliberal considerations involving cheating and information. Grieco argues that both absolute and relative gains must be included in the utility function of states; see Grieco (fn. 14), 40–49.

[70] Snidal (fn. 1), 935, demonstrates that a more powerful state, understood as the actor which is less in need of coordination, will, in a bilateral bargaining situation, secure an outcome closer to its most desired point than its weaker antagonist.

[71] Ruggie, "Continuity and Transformation in the World Polity: Toward a Neorealist Synthesis," *World Politics* 35 (January 1983), 261–85.

Nevertheless, for a very large class of global issues, indeed the classic agenda of the study of international politics—security, autonomy, and the distribution of valued resources—power needs to be given pride of place. These issues are not characterized by the fact that Pareto optimal outcomes are not Nash equilibria. Neoliberal speculations about the positive consequences of greater information are fascinating (even if empirical demonstrations of such benefits are scarce). But they obscure considerations of relative power capabilities, which draw attention to how the payoff matrix was structured in the first place, how the available options are constrained, who can play the game, and, ultimately, who wins and who loses.

SOME CONSEQUENCES OF GIANTISM
The Case of India

By JOHN P. LEWIS

MANY interpretations of India's development experience pay insufficient regard to its size. Diplomats in their formal dealings pretend to be blind to country size. As far as they are concerned, all nations are juridical peers. And economists are almost as bad. They tend to think in terms of ratios and relatives and thereby downplay the effects of country size.[1] Thus, we often speak as if two economies with similar incomes per capita, saving rates, and rates of population and income growth are much the same—even when one is a hundred times the size of the other. They are not, of course. And the difference between the big and little systems is not simply one of calibration. They differ qualitatively precisely because not all of their salient characteristics are proportionate. Some, obviously, are. For example, families, farms, and even factories do not tend, averagely, to be much larger in big countries; the numbers of these entities are, indeed, more or less proportionate to country size. But consider these countercases: the number of a big country's top politico-administrative leaders is a smaller fraction of its population than in a less populous country; the large country's market is proportionate to its size but, thanks to its scale, is qualitatively different; its human resources situation is in some respects qualitatively different

[1] An early conference of the International Economic Association (in Lisbon, 1957) yielded volume of papers edited by Edward Austin Gossage Robinson, *Economic Consequences of ie Size of Nations* (London: Macmillan, 1960). There was thereafter little of a generalized haracter about national size in the economic literature until Hollis Chenery and Moises yrquin, *Patterns of Development* (New York: Oxford University Press, 1975); and Mieko Nishimizu and Sherman Robinson, "Trade Policies and Productivity Change in Semi-ndustrialized Countries," *Journal of Development Economics* 16 (September–October 1984), 77–206. Meanwhile, however, the literature that focused on the small economy, dating from ie Kuznets paper in the Robinson volume above and from William Demas, *The Economics f Development in Small Countries* (Montreal: McGill University Press, 1965), has expanded include, among others, C. Real de Azua, "Small Nations and the 'Constructive' Style of evelopment," *CEPAL Review*, 2d semester, 1977, pp. 153–73; Bimal Jalan, ed., *Problems and olicies of Small Economies* (London: Croom Helm, 1982); Vincente Galbis, "Ministate Econ-mies: Underlying Characteristics and Macroeconomic Policies," *Finance and Development* (June 1984), 36–39; T. N. Srinivasan, "The Costs and Benefits of Being a Small, Remote, land, Landlocked, or Ministate Economy," *World Bank Research Observer* 1 (July 1986), 205–; and DeLisle Worrell, *Small Island Economies: Structure and Performance in the English-eaking Carribbean since 1970* (New York: Praeger, 1987).

from that of a smaller country; and so are elements of its external eco nomic and political relations.

Size effects of the sort just mentioned—on (1) the management of th economic system, (2) human resources matters, and (3) the economy' external relations—are ones this paper seeks to examine in the case o India. First, it may be well to address explicitly the question of size cri teria. This is an issue that has been debated considerably in the smal literature on country size—especially, small country size.[2] There is a inherent ambiguity about whether countries are pieces of geography o populations of people. A continent is clearly the former, a city, obviousl the latter. A nation can be thought of either way, and, further, if it i viewed as a market, size can be said to be a function not just of popula tion but of income per capita. Thus, among national markets, the U.S. i still much the largest.

All of this would be of greater concern if the present piece were pri marily a *comparison* of giants. Primarily, it is a discussion of one big country, India, that is large by any standard. Most obviously, India is large because of its population. Indeed, population seems the dominant measure of national size: nations are most of all collections of people, not pieces of real estate. Planting a national flag over Antarctica would not add that desolate expanse to the world's handful of giant countries.

Accordingly, to the limited extent the following discussion *is* compar ative, the identification of "giants" is simply demographic—and arbi trary. They are the four countries in the world with populations in excess of 250 million, that is, in descending order, China, India, the Soviet Union, and the United States. At the same time, because of the income component in market size, it is often best to consider the relation be tween population size and market size to be a stratified affair; it can be examined in any set of different-sized countries with similar per capita incomes.

I. System Management: Coping with Giantism

The pivotal hypothesis of the present analysis is remarkably simple. It is that a giant system like India's has an inherent, inbuilt tendency to be

[2] Cf. Peter J. Lloyd, *International Trade Problems of Small Countries* (Durham, N.C.: Duke University Press, 1968); Nadim G. Khalif, "Country Size and Trade Concentration," *Journal of Development Studies* 11 (October 1974), 81–85; J. Jainarain, *Trade and Underdevelopment* (Washington, D.C.: Georgetown, Institute of Development, 1976); Jalan (fn. 1); Andrew S. Downes, "On the Statistical Measurement of Smallness: A Principal Component," *Social and Economic Studies* 37 (September 1988), 75–96; and Harold Codrington, "Country Size and Taxation in Developing Countries," *Journal of Development Studies* 25 (July 1989), 508–20.

mbersome. The basic reasons are, first, that most organizations and e activities in which they engage are structured hierarchically and sec- d, that managers tend to have the limited horizontal reach encapsu- ed in the administrative concept "span of control."

There is a cluster of neoclassical economic writings premised on an dividualistic, atomistic state of nature that undertakes to explain why is rational for a work force (taking account of differential skills, etc.) organize itself hierarchically.[3] Other social scientists, being of the view at command systems are about as old as the species, scarcely need such lemonstration. The command structure may be of limited scope, and, in feudal Europe, its writ may not effectively reach the edges of its minal territory. But historically the natural mode of human organi- tion has been hierarchical; the responsibilities (whatever the capacities) political authority have been almost unlimited; and free markets and mocratic modes of governance have had, quite laboriously, to be ved out within the primordial envelope of command structures.

The other factor contributing to the managerial predicament of giant stems is the simple point about span of control —or what is called "span attention" by administrative theorists of a psychological bent. The lit- ature on this point is thin.[4] Although it seems intuitively clear that ere is a narrow limit to the number of subordinates a boss can effec- ely boss, most authors have been reluctant to suggest precisely what is number is. The span of control varies with the work being done (a ill sergeant can supervise a whole platoon) and the particularities of a oduction structure (under quite stylized circumstances a symphony or- estra conductor can manage one hundred musicians directly and effec- ely).

There is some scope, moreover, for stretching an executive's manage-

[3] Cf. Herbert A. Simon, "The Compensation of Executives," *Sociometry* 20 (March 1957), -35; Thomas Mayer, "The Distribution of Abilities and Earnings," *Review of Economics l Statistics* 42 (May 1960), 188–95; Oliver E. Williamson, "Hierarchical Control and Opti- im Firm Size," *Journal of Political Economy* 75 (April 1967), 123–38; G. A. Calvo and Wellisz, "Supervision, Loss of Control and the Optimum Size of the Firm," *Journal of Polit- l Economy* 56 (October 1978), 943–52; idem, "Hierarchy, Ability and Income Distribution," *rnal of Political Economy* 87 (October 1979), 991–1010; Sherwin Rosen, "Authority, Con- l and the Distribution of Earnings," *Bell Journal of Economics* 13 (Autumn 1982), 311–23; j Kumar Sah and Joseph E. Stiglitz, "The Theory of Economic Organizations: Human libility and Economic Organization." *American Economic Review* 75 (May 1985), 292–97.

[4] According to Lyndall Urwick (who drew on Henri Fayol, the French pioneer in admin- ative theory, and who collaborated with Luther Gulick in putting out in 1937 *Papers on Science of Administration* as a contribution to the study of reform of the U.S. presidency which Gulick was engaged), "No supervisor can supervise directly the work of more than e, or at most, six subordinates whose work interlocks; Urwick," *Elements of Administration* ew York: Harper, 1943), cited in Bertram M. Gross, *The Managing of Organizations* (Glen- , Ill.: Free Press, 1964), 148.

ment span through the use of assistants and technological reinforcements. But none of this disturbs the proposition that in each given organizational and programmatic context the average number of subordinates that superiors can effectively supervise is finite and small. Further, there is little evidence that these manageable limits of supervision are much larger in big systems than in little ones.

The result is inescapable: large systems have more bureaucratic layers. They also have wider hierarchies. Even if spans of control are uniform, there are more vertical chains of command and communication that invite coordination as one adds lower levels to the pyramid and views it cross-sectionally. Moreover, there is incentive in the giant system to contain the layering, which leads to a broadening of average managerial spans and therefore widens the bureaucratic structure still more.

Bertram Gross, in his monumental work, *The Managing of Organizations*, remarks, "I have found very few organizations with 10—and none with more than 12—hierarchical levels (as contrasted with the much greater number of levels often found on salary scales)."[5] However, the number of people Gross has in mind refers simply to those within the organization itself. Thus (using Urwick's control span of 6), he notes that a mere nine tiers can encompass a work force of two million. For a discrete institution—say, an automobile manufacturer or a foreign affairs ministry spread worldwide—that already sounds very large, and, as remarked already, most discrete organizations (that is, families, farms, and so forth) are not much larger in big countries than they are in small ones. There *are* a few instances where the in-house scale of an institution does tend to be proportionate to national size. One obvious case is the army (although the cross-country fit of this generalization will not be very close), and here a bigger military, whether of one country over another, or in the same country over time, typically means more layers. This presumably was the reason for the creation of five-star U.S. generals and admirals in World War II; naval personnel, for example, increased twelvefold.

The more important point that Gross's reflection seems to miss, however, is that many public bureaucracies—agricultural extension services, postal services, tax-collecting agencies, welfare systems, public schools—do business with citizen clienteles. These clienteles in India are ten times as large as in a country of eighty million and a hundred times those of a state of eight million. Therefore, unless teacher-pupil ratios and the other comparable ratios at the business end of these service and delivery

[5] Gross (fn. 4), 375.

systems are notably larger in India than in smaller countries, India has an extra bureaucratic layer or two—as well as, by the same token, a broader horizontal span of hierarchy.

The final point for establishing the system-management problem with /hich giantism must cope is that "another layer or two" of bureaucratic ıyering is no small thing. Extra layers inevitably impede operational ommunications—and, therefore, operations themselves. Extra layering arbles, distorts, delays, and loses commands, responses, and other mes- ıges being passed up and down the hierarchy, as well as laterally across . The operations of giant high-rise bureaucratic pyramids have an in- erent disposition to be sluggish and clumsy.

Furthermore, using Gross's arithmetic and assuming a span of control f as high as 10 at the lower levels of our hierarchy: to add, say, a tenth nd eleventh level to the bureaucracy in order to go from a country of 8 ıillion to one of 800 million one does not merely add (arithmetically) 10 r 20 percent to the noise and static in the system. There is an exponential ıcrease, as distortions, gaps, delays, and misunderstandings are com- ounded up and down the line and horizontally as well.

There is room for far more empirical research into these matters of rganizational architecture, as well as for efforts to collate the several ılklores concerning hierarchy and authority spans that pertain in differ- nt kinds of institutions—religious, educational, corporate, military, overnmental, and other. Meanwhile, however, the proposition that pub- c bureaucracies serving giant clienteles tend to be cumbersome seems) be nearly self-evident. Most leaders of the giant bureaucracies are ware of the problem and seek ways to cope with it. The Indian case can e so interpreted; and although the matter will not be pursued here, ecent commentary extends the same perspective to the other three gi- nts, China, the USSR, and the U.S.

However political and bureaucratic leaders try to temporize with the wkwardness of giantism, there are only two real strategies for coping 'ith it. One is to decentralize by delegating (insofar as this consistent 'ith the system's collective purposes) categories of decision making to ›wer levels of the hierarchy. The other can be called sideways delega- on: governments cannot escape their ultimate responsibilities; but 'ithin sectors for which they are responsible they sometimes can find ıbor-saving devices, that is, largely self-adjusting servomechanisms that an generate streams of socially acceptable decisions without requiring ontinuing day-to-day intervention by officials. The most familiar and nportant such mechanisms are markets populated by many buyers and ıany sellers, each with considerable operating discretion. But some

other collectives and networks of nonofficial institutions are perceived to have similar servomechanical properties.

DECENTRALIZATION—IN GENERAL AND IN INDIA

In important respects decentralization runs against the human grain. Political and bureaucratic leaders on average have a revealed preference for power; they tend to share it only grudgingly. And many of them value the rents that come their way and the way of their in-groups in the course of centralized decision making. These related motivations have inhibited decentralization in India, as we shall note.

Nevertheless, it is also human to wish to reduce one's workload when this can be done without much sacrifice of prestige or emoluments. Thus, leaders' self-serving preferences as to decentralization are offsetting. There is room for coping strategies motivated by what the leaders perceive to be the public interest.

It is recognized that the downward delegation of decision making in a hierarchy (or outward delegation in center-periphery terms) must avoid unacceptable losses of coherence. Thus, national leaders are committed to a set of national goals. The latter, which may be comparatively few, exist alongside all the limited goals being pursued within the system at personal, household, enterprise, local, and subnational regional levels. The local goals should not be, and typically are not, inconsistent with the national ones, but the same could be said of many alternative micro purposes and efforts. Further, certain subnational efforts demanded by the national objectives should be made, and quarrels and differences between units of the hierarchy, including those arrayed in parallel, should not be counterproductive.[6]

Within these constraints, there are two principal substrategies for safe decentralization.[7] One is that of indoctrination. Most bureaucracies engage in some of this. Officers down the line are trained and conditioned to act, indeed, to think, in the mode of a particular company, service, agency, church, or government. There are cases of superindoctrination that rely almost entirely on this approach. Before the day of electronic communications, this had to be the way in which naval establishments achieved desired and predictable performance from their commanders

[6] Not all internal differences, even conflicts, are to be avoided. Competition among collateral units is usually healthy, and many systems place little premium on lockstep conformity.

[7] The constraints do not apply to a third case, where the performance of the center's objectives is not measurable. Thus, for example, if the sole purpose of an organization is to save souls, the implementation of the activity can be radically decentralized. But if the success of joint efforts is subject to common observation, those in whom responsibility for the common effort is centered must limit themselves to some combination of the two alternatives in the text.

t sea. Similarly, before World War I the officer personnel of the impe-ial German army, more particularly of the general staff, were selected nd indoctrinated to become intellectually interchangeable with any ther member of the staff—to the point that a junior staff officer could nange his superior's order (because of the junior's better on-the-scene nformation) and then, on behalf of his superior, give orders to a field ommander of higher rank than himself, which the latter was obligated o obey.[8]

It is worth noting, for reference below, that decentralization via in-octrination tends almost inherently to be elitist. It involves the creation f a special service of agents, individuals who can be entrusted to do the ill of the center. In most bureaucratic instances, however, indoctrina-on is only a supporting strategy. The main decentralization scenario is ne that was articulated for India by two economists, D. R. Gadgil and ukhamoy Chakravarty, who in 1972 succeeded each other as members f the national Planning Commission and, in succession, took an interest n promoting a proper system of "multilevel planning." Clearly, they aid, there needed to be more downward delegation in official economic ecision making, and there was no virtue in uniformity for uniformity's ake. If, on the one hand, a decision concerning a local matter made little ifference to neighboring jurisdictions or levels above, let it be settled ocally. If, on the other hand, local activity could contribute to goals being ought up the line (say, a local choice to grow grain rather than fiber ould help the country's drive for food self-sufficiency) or if local activ-ies could have significant negative or positive effects externally (for ex-mple, on the rest of a watershed lying beyond this locality), the decision hould be escalated to the first higher level within which it would fit ithout spilling significant externalities outside.[9]

Although prodecentralization doctrine of this sort has found official vor in India for some time, its implementation has nonetheless been esitant and halting. This has been true of both zones—macro and mi-ro—into which questions of downward delegation divide.

Like the other demographic giants, India has broken its official hier-chy into two overlapping levels—"Centre" and, in its case, states (in

[8] *Encyclopaedia Britannica*, 13th ed. (1926), s.v. "Staff." (The first 28 volumes of this edition ere identical to the 11th edition of 1910–11; three supplementary volumes were added to over the period 1910–26.)

[9] A reader of an earlier draft of this paper objected that the rule in the text does not cover e case of a national government that may appropriately, as a matter of policy, wish to retain ntrol over a local activity (say a toxic one) even if it imposes no burdens or benefits beyond e local jurisdiction. I would argue that the bad precedent or damage to the national body olitic that the denial of local option would forestall does indeed fall into the category of egative externalities—as could the abandonment of various national cosmetic uniformities at orderly centrists might value.

China, provinces, in the Soviet Union, republics, in the U.S., states). India's federalism, like America's, is formalized in its constitution. The states have a measure of autonomy. Questions of macro delegation concern the allocation of authority and duties between Centre and states.

Here the situation can be summarized fairly quickly. The states' constitutional autonomy is real but very limited; the Centre has greater opportunity than in most federal systems to dismiss state governments, and it has exercised the power repeatedly. Partly through the mechanism of five-year planning the Centre's interventions in economic policy have extended regularly and heavily into such "state subjects" as agriculture; and the states have been overwhelmingly financially dependent on the Centre, especially for public investment and other developmental outlays.

Yet it would be a mistake to see the Indian states of the 1960s as pawns or ciphers. Interest groups arrayed themselves at the state level. State legislation was consequential. States balkanized the food map; and they battled fiercely over such things as public sector factory locations. What kept Centre-state relations in reasonable balance during the first twenty years of independence was the unbroken hegemony of the Congress Party, not only nationally, but in most of the states, and the fact that there was Centre-state balance in the party itself, the governing councils of which were dominated by independently powerful regional chieftains.

Things changed from the late sixties onward: non-Congress, including regional-party, governments became commonplace at the state level. In the Congress power and finance (including, by common perception, illicit finance) were visibly concentrated at the Centre around the person of the prime minister, who conducted an authoritarian experiment between 1975 and 1977 that further agitated politics within the states. Corruption became more visible in several states, as did ethnic conflict and state chauvinism. At the same time, the diverse economic fortunes of the different states probably augmented welfare disparities between states that had been accruing since independence. More recently, moves toward economic liberalization have made more feasible the disaggregation of some governmental economic policies and operations. This together with the greater salience of regional parties and politics made manifest in the election of 1989 may well augur a greater devolution of economic decision to the states—with both a promise of faster overall growth and a risk of aggravated balkanization.

Meanwhile within modern India the more pervasive, if less conspicuous, centralization-decentralization issue has been the micro one, carried over from British times: *Within* states how much decision making should

be made locally? And by whom—the locals or resident representatives of the Raj? Before and after independence the Indian administration opted for the superindoctrination solution. The country was governed with a remarkably thin civilian officer corps that, although composed of state cadres, was an elite all-India service whose governance extended all the way down to the subdistrict level. Before independence the Indian Civil Service was progressively Indianized; following independence the Indian Administrative Service continued in the same frame. It placed remarkable discretion in the hands of thirty-year-old district "collectors" (or magistrates or administrators), who, having been carefully selected and trained, were trusted to make the same decisions their superiors would have made had they been there. By the same token the system placed precious little authority in the hands of local leaders and electors.

As representative democracy developed at the state level, members of (state) legislative assemblies in a number of states exercised considerable influence in their resident areas and, in effect, forced the incumbent IAS officials to share power with them. But neither of these elements had great appetite for local self-government. Effective partisanship for the latter, as, indeed, for decentralization in general, tended to come (sparingly) from a few publicly interested national and state leaders. Thus, at the end of the 1950s, prodded by official reviews that found the participatory aspects of the Community Development Movement inadequate, Nehru sponsored the establishment of the reforms and structures of *panchayati raj*.[10]

But in the early years these were taken up in a serious way only in two states, Gujarat and Maharastra. During the later sixties and the seventies a number of states stopped holding local panchayat elections altogether. More recently, there has been renewed and vigorous experimentation with local self-government in some states, notably West Bengal and Karnataka, and as of mid-1990 the disposition of the postelection Union Planning Commission indicated a possibility of new momentum for microdecentralization.[11]

[10] *Panchayati raj* means literally "government by local councils." The recommendations were those of the Balwantrai Mehta Committee in 1958, which proposed a three-tier local government structure—village panchayats, panchayat samatis (societies of panchayats) at the block level, and district zilla parishads (district committees). See, for example, Marcus Franda, *Small Is Politics: Organizational Alternatives in India's Rural Development* (New Delhi: Wiley Eastern, 1979), chap. 4.

[11] While this is not the place to review the extensive literature that has developed over matters of local institutional reform in India, let me note two of my own writings on the subject: "Notes of a Rural Area Development Tourist," *Economic and Political Weekly*, June 29, 1974, pp. A42–54; and "Overview" in John P. Lewis, ed., *Strengthening the Poor: What Have We Learned?* Overseas Development Council Policy Perspectives No. 10 (New Brunswick: Transaction Books, 1988).

It is not surprising that in India an augmented interest in decentrali zation translated immediately into a concern for decentralized, that i: "multilevel," planning. As with the movement toward downward dele gation of official decision making generally, there has been some ten dency to confuse cosmetic with genuine planning at the local level. Man instances of so-called district planning, for example, have been done, nc by planners who belong to the districts, but by professionals of the stat planning agency deputized to plan *for* the district. Similarly, efforts t staff local planning undertakings with appropriately trained locals hav been spotty. At the same time conceptions of how extensive and sophis ticated local planning exercises need to be often get greatly exaggeratec There is no need for "mini-macro" planning of the same scope and com plexity as that which the Union Planning Commission does nationall What is needed is some sensible short- and medium-term expenditur budgeting for those public sector activities that local authorities are com petent to undertake, together with a plan for their financing (includin projections of grants in aid from higher authorities) as well as a plan fo the selective development of market-town type, rural-urban centers.[12]

Markets and Other Servomechanisms

Markets enter our story both defensively (as instruments of sideways del egation, taking some of the burden off public bureaucracies) and posi tively (bigger markets are often alleged to be better ones). The case for a positive size gradient in other nonofficial servomechanisms (for example cooperatives and private education networks) is less clear, but they toc can share the bureaucratic burden.

In the 1960s many inside as well as outside the government of Indi began to wonder whether much of the detailed economic decision mak ing that government functionaries had been doing directly (for example allocating imports, choosing investments, and setting input and outpu prices) would not be better left to the various markets. The latter could be allowed to adjust themselves competitively within an environment se

[12] For an elaboration of the local-planning points, see Lewis (fn. 11, 1974); also V. M. Rao, "Decentralised Planning: Priority Economic Issues," *Economic and Political Weekly* 24, no. 25, June 24, 1989, pp. 1399–1405. On the spatial aspects of decentralized development, poten tials for selective small-town development are examined in John P. Lewis, *Quiet Crisis in India* (Washington: Brookings, 1962), chap. 7; and E. A. J. Johnson, *The Organization of Space in Developing Countries* (Cambridge: Harvard University Press, 1970). The fear of metropol itan bloat expressed in these early works has been found to be exaggerated by Rakesh Mohan and Chandrasekhar Pant, "Morphology of Urbanization in India: Some Results from the 1981 Census," *Economic and Political Weekly*, September 8, 1982, pp. 1534–40, and September 15, 1982, pp. 1579–88.

by broad commercial, monetary, and fiscal policies and framed by government regulations designed to permit as much freedom to transactors as was consistent with the avoidance of cheating and unreasonable exercises of market power. Then and much of the time since, the tilt of official opinion in India, like the views of such multilateral agencies as the International Monetary Fund (IMF) and the World Bank, has been toward such economic liberalization. Many prefer to hoard their power and rents and therefore resist such a policy agenda, but even so it is widely agreed that freer, albeit bounded, markets can improve the efficiency of the system. While liberalization does not promise to yield equitable distributions of rewards among individuals, classes, sectors, and regions, such reform can at least eliminate some of the inequities generated by the control systems it displaces. As to redistribution, liberalization can be supplemented by fiscal policy and by direct attacks on poverty, including transfers to the poor of such assets as land and, by means of education and training, new human capital; all of these measures can be undergirded by strategies seeking to empower the poor politically.

Markets are the principal but not the only systems that can run themselves in benign ways without constant government steering. Elaborate arrays of private schools or other educational establishments often conduct and finance themselves without extensive public funding or control. The same is true of many private health institutions and programs. Systems of cooperatives can be, and in some Indian cases have been, useful, largely self-managed supplements to the private and public sectors between which they stand. Networks of private philanthropies and nongovernmental nonprofit organizations can provide interesting and valuable augmentation and/or substitution for a variety of public programs. Many professions largely regulate themselves.

It would be difficult to establish that giant economies have had more resort than medium or small ones to these various servomechanisms. But, according to the present reasoning, their *need* for sideways delegation is unsurpassed.

Development economists have asserted that a bigger national market is also better and that this is the principal advantage of large national size. They have recognized that in a large country the obstacles to internal integration that are interposed, in part, by the terrain and the costs of internal communications may leave the potential of the "national market" unrealized. Yet the national frontiers are the ones that establish the limits of the main regulatory regimes that order transactions. They delineate currency and financial regimes, and they commonly raise tariff

and other obstacles to transactions across the frontiers. In short, nationa markets are consequential, and there is a commonly held view that thei size and quality are positively related:

—Large national markets are characterized by a greater and finer division of labor (more specialization) and therefore by greater gains from (*internal*) trade.

—Insofar as development entails the introduction of new productive processes and products into poor countries, processes enjoying economies of scale and depending on internal demand can fit earlier into a large poor national market than into a small poor market.

—Thus, where there are economies of scale, large markets can accommodate a proliferation of efficient-size firms and thereby gain the benefits of internal competition sooner than small markets can.

—If the constraint on the introduction of a new lumpy investment is the availability of savings, and one imagines a large country and a small one with equally good financial intermediaries, those in the large country should be able to skim the needed absolute capital from a thin flow of savings sooner than their smaller counterparts.

To be sure, there are major challenges to the linkage of national siz and market quality. First, a good neoclassical economist[13] will argue tha a small economy can enjoy all the alleged advantages of a large domesti market if it simply has access to an open international market. There car be specialization and division of labor across frontiers. Hence, new ca pacities suited to the comparative advantages of the smaller economy dc not need to "fit" within the internal market, and the discipline of com petition can come from abroad as well if there are no protectionist obstructions. So can the savings for lumpy investments that genuinely promise to be rewarding.

There is the practical point that it is hard for the giant economy (which, as will be explained, is altogether likely to have a low ratio of external trade to economic activity) to turn inward without becoming insular. In a large continental, domestically oriented national system, the allocation of the nervous energies of leaders and entrepreneurs to external economic matters tends to be meager. Given their preoccupations with the spacious internal market, it is hard to gear up external marketing, including the design of products for the export market, to world-class standing. As in the United States, the intensity of the outreach remains inferior on average to that of the Koreas, Taiwans, Japans, and Germanys—and, in the last century, the U.K.'s—of this world.[14]

[13] The following is the gist of Srinivasan (fn. 1).

[14] A related consideration is worth noting. For efficiency the giant system depends more than the small country on the discipline of *internal* competition. Thus, it is a mistake in India,

The Indian record also suggests that giant size can increase the system's capacity for making mistakes. India typically is pilloried for the extent to which, in the fifties, it chose the path of self-reliant, export-pessimistic, import-substituting, capital-intensive development. But that India led the way along this path was not because P. C. Mahalanobis had a unique ability to misguide.[15] Nearly all the non-Marxist new states of that period were similarly disposed; India simply went farther than the others because its size made balanced-growth, import-substitution choices more plausible.

A final challenge to the linkage of large economy and market quality concerns technological dynamism, and it is a worrisome one. While the efficiency of the Indian system can and should be entrusted mainly to vigorous internal competition, this may not be enough to ensure that the country becomes as technically progressive as the government and the public desire. Ideally, because of its factor proportions, India needs different (in particular, more capital-saving, labor-using) brands of technology than those most prevalent in much of the world. There is thus great need for indigenous R & D that is innovative as well as adaptive. As will be noted, India already has large cadres of scientific and technical personnel, but these are not yet, in the aggregate, very effectively engaged in the generation of relevant, total-factor-productivity enhancing R & D results. There are moves afoot, however, to invest more effort and resources in indigenous innovation.

In one sense, though, India's technological needs can probably never be met wholly from internal sources, at least not in the longish medium term. Despite the current inappropriateness for India of much of what is going on in global R & D, Indian R & D needs to remain constantly and closely tapped into the world's global centers of research and innovation. But this is difficult for a giant inward-turned system that is still in a fairly early stage of economic development. To do it easily, the giant poor country would need to draw some of these current wellsprings of global research and applied research into itself—and that is almost impossible. Among the giant countries, the United States was uniquely fortunate in this regard. Even in its early stages of development it was never poor in the sense of income per capita. It was the recipient of not only the largest but also the most talented brain drain in the world's history—which, of

---

as it is in the United States (the opposite cases could be Taiwan and the Netherlands), to regard antitrust policy, or procompetition regulation, as a form of interference with market outcomes that a proper liberalization program should scrap, along with other forms of trade, investment, and price control.

[15] Mahalanobis was the principal designer of India's Second Five-Year Plan, 1956–61.

course, continues, in part from India. Thus the U.S. had the best of both alternatives: it could look inward technologically and still be on the innovative and intellectual frontiers. It quickly became the center of global innovative activity (although the claim now is being challenged by Japan and others), with the periphery having to come to it.

India has no real chance of following this model. If it nevertheless follows the instincts of a giant—giantism makes chauvinism more plausible—it is likely to become a huge, rather backward, somewhat second-rate, but quite self-reliant recluse. There are some symptoms of this. One thinks, for example, of the din over the years that the indigenous fertilizer research unit at Sindri has been putting up against successive foreign technological encroachments. While recognizing the excellence of many Indian scientists and engineers, one considers what large fractions of India's vaunted armies of technical personnel are in fact poorly trained because the many educational and training institutions are ill equipped to cope with the demands for instruction and credentialing being placed upon them. One remembers ministers throwing up their hands over the impossibility of keeping up with the pace at which various global technological developments were moving. One thinks, indeed, of the Hindustan Ambassador automobile. Surely there is no other country in the world, at least no pluralistic one, in which an already retired Morris Oxford of the 1950s could, as the joint product of government policy and consumer stoicism, have remained the dominant passenger car model for three decades. There is much in this particular symbol that is attractive—some asceticism and the self-confidence of a system large enough to set its own norms. But it is not the mark of a country bent on technological dynamism.

The short of it is that if India does not wish to remain a soft, technically lagging economy, it must promote technological intercourse with the rest of the world. Partly this can be achieved by continued movement of people and ideas. But a good part of it requires imports of products. Even if vigorous internal competition can do most of the proefficiency disciplining in a giant market, competition from imports is needed for technological invigoration. Plainly, moreover, if more imports are to be consistent with progress toward financial self-reliance, an increase in imports will require more successful expansion of exports. Here is one of the places where policy must try hardest to override the natural gaiting of a giant pluralistic system.

Another carrier of technological interchange, however, can be foreign investment, especially direct investment by foreign enterprises, and here the giant host country has certain natural advantages if its policies do not

suppress them. First, the giant host has less to fear; the danger of its being dominated by any multinational company or cluster of multinationals is far less than it is for a little country. Second, its internal market, compared with those of smaller countries of the same income level, is of greater inherent interest to foreign investors.

Markets: Some Conclusions

The national market of the giant economy can be a net resource if it is safeguarded. The neoclassical alternative of a completely open international market is simply not available. The dangers and disadvantages of a giant internal market are subject to domestic policy correction. Thus, the next market consequences of large size should be positive—if Indian governments in the years ahead can adopt policies for pushing exports somewhat more effectively than would be the collective inward-looking inclination of local entrepreneurs left to their own devices; if future governments also can exploit the system's potential for attracting constructive foreign investments; and if government can sustain an efficient regime of procompetition policy in the internal market.

Contrary to such a prescription, Indian governments at all levels have a long history of shooting themselves in the foot. It has been their custom to obstruct the internal market. This is the effect of every octroi that local jurisdictions assess on goods traffic to and through their territories. Lately some of the states have been trying to block or discourage the in-migration of certain categories of the nation's citizens and/or residents. Best known and perhaps most distorting are the various barriers, including food zones, that have been erected to food trade, sometimes at the instance of states or of state politicians but always at least through the sufferance of the Centre.

There is no way a balkanized internal market can reap the benefits of large national size. Internal obstructions are not necessarily inane. The octrois might make sense if no other tax base were available, and, in the era of famines when the ultimate responsibility of a provincial government was to assure food supplies for its people, restrictions on food movements were a means to that end. Nevertheless, such barriers inevitably impair the efficiency and dampen the growth of the national economy.

II. Human Resources Dimensions

Big countries often are alleged to be peculiarly burdened with human resources problems that in fact are no more than proportionate to their

size. Thus, although China and India are spoken of as if the corner on "the" population problem (that is, the populatioı problem), population growth in those two regions actually poses of a problem than it would if they were both subdivided into tl ponent provinces. Plainly, India, with an annual population g little more than 2 percent, not to speak of China, with a rate c percent, does not face nearly the demographic threat that ⁄ growth poses for Kenya.

Similarly, although Arthur Lewis says he had India in mind wrote "Economic Development with Unlimited Supplies of L labor redundancy is not a function of total population or size of force; it reflects an abundance of labor relative to other produ tors, particularly capital. Thus, in South Asia, India may not fit t paradigm as well as does Bangladesh with only one-seventh the tion or even Nepal with only one-fiftieth as many people.

Personnel: Training and Critical Mass

But then, as to human resources, where *do* the proportionaliti down? For one thing, there are instances (notably with respect ı armies) where the absolute number of persons a government c lize is important. There are also instances where the larger recruitment of key personnel can be more selective. If one assu (1) talent is randomly dispersed through all populations, (2) a able to attract some of its best talent to political-administrative and (3) its selection procedures are such that it in fact chooses and brightest applicants, then, theoretically (if these assumptic India), incumbents in the country's upper political-administrat lons would be more gifted on average than those in a small cou

That specific proposition need not be pressed very seriously principle—that if one is seeking an absolute number of perso average quality of the complement that can be skimmed fron population with a normal distribution of talent usually exceeds be skimmed from a smaller population—is persuasive. So ar size-related assertions that can be made about training: ex worldwide makes it quite clear that training programs for themselves best organized and conducted along elitist lines. middle if not early educational stage, the most promising ca need to be gathered into selective centers of excellence. But

[16] This he says in conversation, not in the article itself, which appeared in Development with Unlimited Supplies of Labour," *Manchester School* 22, 21 (May 91.

principle about training elites also is clear. The quality of training is in part a function of the size of the centers and their programs. Excessive size may diminish quality (although this is less clear if there is scaling up all around and the quality of the trainees is not diminished). But, very clearly, excessively *small* size is an enemy of quality. If programs have too few participants they may not be able to afford adequate laboratories and libraries or provide as specialized and subdivided a curriculum as state-of-the-art training requires. Larger student bodies tend to be more competitive. And even if the quality of individual faculty members and the faculty-student ratios remain constant, there is inherently more opportunity for interpersonal interchange—and more variety in the interchange—as the size of small programs increases.

If it is true that the best students in the best programs get a good part of their teaching from each other and if there is as well something to the high value commonly placed on the productivity of collegial interchange among researchers and teachers, then a training institution's potential for one-on-one interfaces is one of its major assets.[17] On this count sizable programs beat miniprograms hands down. Just as a giant country can accommodate a larger number of manufacturers of efficient scale in its domestic market, it can accommodate a larger and more diverse array of elite training establishments of minimum effective size. The experience of India in this regard compared with that of Bangladesh and Pakistan provides rough but robust support for this proposition.

Moreover, trade in this case is not as good a substitute for large size as it is in the case of commodities and other products. Indeed, in the absence of indigenous training for elites, smaller countries import the training; that is, they export the trainees. This probably increases the country's exposure to brain drain. Moreover, the smaller country getting more of its training abroad is also slower to form viable indigenous centers of excellence around which critical masses of expertise can gather.

The concept of critical masses of expertise deserves some attention. The numbers in a professional or paraprofessional cadre can be said to have reached a critical mass (regardless of whether some of the final training is still being done overseas) when the cadre has developed a distinct self-awareness and a credentialing process; when it is internally networked and has mechanisms for establishing superior professional status; when members have recognized specializations; when they are

[17] The number of potential one-on-one interfaces (N) among individual members of a group rises exponentially as the number of members of the group (a) rises. Specifically, $N = \frac{a(a-1)}{2}$.

afforded a full professional life and are enlisting new recruits at least a
a replacement rate; and when, in particular, members' net migratio
abroad has not, from the viewpoint of the nation's interest, become dys
functional.

Such critical mass is much likelier to be obtained when numbers ar
large. A large cadre with a greater assortment of members is likelier t
have one who can be, if not an interchangeable part, then at least a ser
viceable replacement for another pivotal member who departs. It has
much larger number of professional interfaces. And an absolute numbe
of out-migrants is a smaller relative loss for a large cadre. Moreover,
large cadre has a better chance of generating and sustaining sufficien
pioneering and ongoing professional activity at home to become a mag
net for some fraction of the profession's numbers sojourning abroad.

The Indian economics profession is a case in point. In 1975 I partici
pated in a comparative study of the development of economics in India
Pakistan, and Bangladesh.[18] At that time we determined that, of all th
Pakistanis with overseas postgraduate degrees, more were then em
ployed in the World Bank and IMF jointly than in all Pakistani univer
sities. The profession in Bangladesh was being similarly decimated b
emigration. In India, by comparison, brain drain was no problem—eve
though in absolute terms there was more of it and the quality of some o
the emigrants was extraordinarily high. The emigrants were not reall
that much missed. Their achievements continued to be identified wit
India. They continued to visit. The production of new first-class India
economists continued apace, both at home and overseas. Most of the ke
chairs in government and in the policy analysis and academic commu
nities were occupied by able persons, and people kept returning from th
World Bank and elsewhere to government and institute jobs at home
lured by the attractiveness of the work and colleagues and despite th
financial disincentives.

Not all Indian professions and disciplines are so well positioned ir
these regards, but the condition is not uncommon. It would appear tc
extend, for example, to much of engineering, the natural sciences, anc
medicine. And where critical mass prevails, the disadvantages of brain
drain may be greatly reduced. It still can be the case that, as the benefi
ciaries of publicly subsidized education, highly trained emigrants are the
bearers of reverse foreign aid to such places as the United States. But this
would be true also for a small exporting country and small exporting
professions. Where there is a critical mass, the losses emigration imposes

[18] Prepared for the Ford Foundation, 1975.

›n the system are less. Indeed it is possible that the emigrants' remit-ances exceed what their social products at home would have been. -Iigher education is such a labor-intensive industry that it could be to ndia's continuing comparative advantage to export a considerable vari-:ty of trained professionals.

III. Giantism Internationally

Гhe Propensity to Trade

'τ has been widely assumed that low ratios of international trade to na-ional product in the demographic giants constitute clear evidence of ʌrongheaded, that is, excessively inward-oriented, policies. Indeed, in he past India probably overdid import substitution. But the lowness of ts ratio of international trade (exports plus imports) to gross domestic ›roduct—the figure was 14 percent for both India and the United States ı 1986[19]—can be attributed mainly to the size of the Indian economy.

This last makes a priori sense. Larger systems have greater diversities nd scales of internal trade, and if smaller systems are joined together ıto a larger one, the latter is bound to have a lower trade-to-gross do-ıestic product (GDP) ratio than do any of its components. Much of the omponents' previously external trade will now be internalized. Simi-ırly, if the Punjab were unstitched from India, its ratio of trade to GDP ⁄ould, by definition, be far higher than India's before the hypothetical artition.

The proposition that a country's propensity to trade is in part a func-on of national size also can be established empirically. The 1988 *World ꞌevelopment Report*'s basic tables provide merchandise trade and GDP ata on 100 countries whose populations ranged in size in the early 1980s 'om about 1 million to China's 1,019 million. Their trade-to-GDP ratios aried from Singapore's 277 percent in 1986 to Burma's 11 percent. An ıspection of the array suggests that the propensity to trade internation-llу tends, across the set, to be lowered by two factors: large population ze and low per capita income.

Ten countries in the group of 100[20] had a population that exceeded 70 ıillion in the early 1980s. Of these, four (India, the U.S., Brazil, and uite strikingly, Japan) had trade-to-GDP ratios of less than 20 percent.

[19] The estimates in this section are drawn from the merchandise trade and GDP series in ıe basic tables of the *World Development Report 1988*, published for the World Bank (New 'ork: Oxford University Press, 1988).

[20] I have left out of the present set the high-income oil-exporting countries as well as those, ıch as the Soviet Union, for which the 1988 *World Development Report* does not provide 'ade and/or GDP estimates.

China would have fallen into this group two years earlier, but in 198(downward revision of GDP estimates together with ballooned impo raised its trade ratio to 27 percent. Four of the other ten largest countr (Bangladesh, Nigeria, Pakistan, and Mexico) had trade ratios of less th 30 percent. The only one of the ten largest with a trade ratio above t 30 percent mark was Indonesia, with 38 percent.

The association of poverty with a low propensity for trading becom evident when the 100 countries are grouped by income levels. The 2 2 matrices shown in Figure 1 are for four groupings: 25 low-incor countries, 33 lower-middle-income countries, 23 upper-middle-incor countries, and 19 industrialized market economies (the better-off mei bers of the Organization for Economic Cooperation and Developmen The partitions between the cells are the two just mentioned: a populati of 70 million and a trade-to-GDP ratio of 30 percent.

Among countries other than the largest ten, it is apparent that t relative incidence of low trading propensities is much greater among t poorer countries. Standing this observation on its head, an advocate outward orientation might well suggest that one reason for the pove of poor countries is their weak articulation with the international ma ket. As we have seen, the same point has some merit in the Indian ca: Yet, decomposed along the present lines, the 100-country array of tra ing ratios makes the strong association between large national size and low propensity to trade very clear.

Giant size does offer an opportunity for closer and earlier approach national self-supply. At the same time, as noted, the giant must gua against the considerable risks of technological chauvinism—and, in th cause, struggle to export and welcome foreign investment a degree mo ardently than would be its natural disposition.

Multilateral Governance and Foreign Aid

In a world defined by sovereign states, governments and their multila eral creatures tend, in their external relations, to address and count oth governments, not people or other entities. This syndrome of countir flags rather than noses diminishes the representation of big countries multilateral councils. The double-deck composition of the Unit Nations Security Council does give the United States, China, and tl Soviet Union, along with France and Britain, a stronger voice in U geopolitical deliberations. But in the General Assembly and in the (noi Bretton Woods) United Nations system generally the one-flag, one-vo rule prevails. It gives India, like the other giants, the same formal clo as the Seychelles. This bizarre allocation of political voice detracts fro

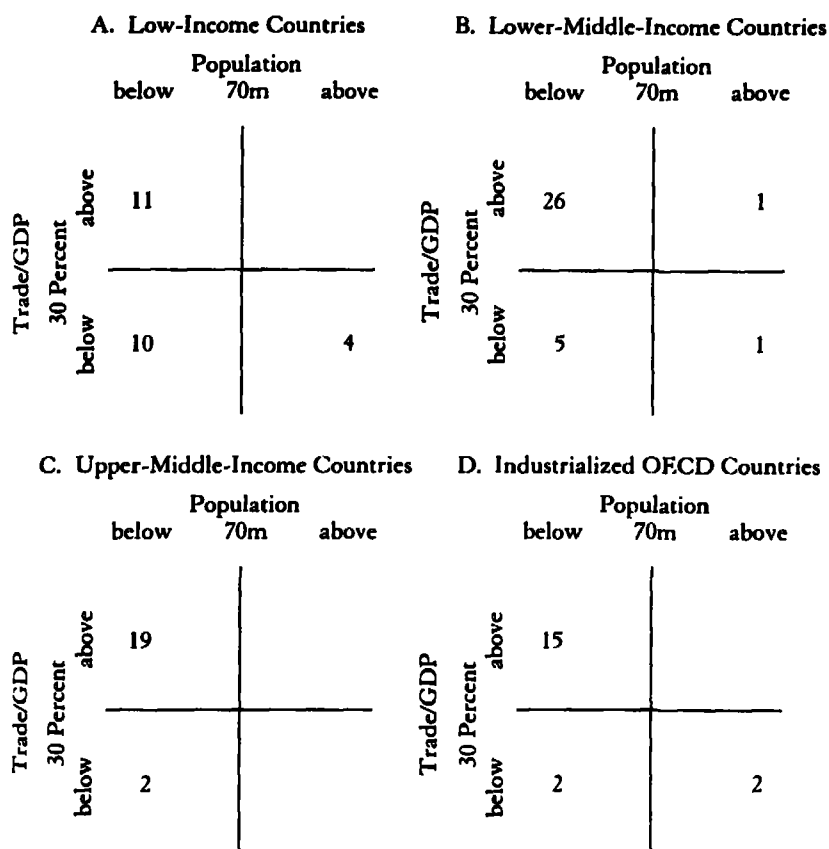



FIGURE 1
100 COUNTRIES IN 1987 GROUPED BY SIZE, INCOME, AND PROPENSITY TO TRADE[a]

SOURCE: The World Bank.

[a] Data and income-level classifications from the World Bank's *World Development Report 1988* (New York: Oxford University Press, 1988). "Trade" includes imports plus exports.

the seriousness with which the larger OECD countries and, especially, the United States view the United Nations and its constituent bodies. The same discount may not apply in the case of the two developing giants of Asia because of the strength of their identification with the North-South bargaining and debates conducted in UN forums. Nevertheless, measured against the one-man, one-vote norm now accepted within most

nations, India's underrepresentation at the United Nations plainly i worse than that of the United States.

The Bretton Woods institutions and the regional development bank patterned after them are a different story. They weight voting system approximately by countries' aggregate incomes. This is a more aggree able arrangement for the larger OECD countries, and, indeed, it gives mor formal representation to China and India than they have at the UN China ranked number 7 and was the first developing country on th World Bank's list of real 1986 GDPs. India, following Spain and Brazil was number 10. (The Soviet Union was not on the list.) Nevertheless under the Bretton Woods dispensation the two Asian giants still fall fa short of the votes they would have under representation weighted by population.

The other manifestation of the squeeze that the multilateral system—and in this case, also, programs of bilateral development assistance—place on the larger poor countries is financial. Aid agencies and devel opment banks disburse their transfers, concessional and otherwise through an array of country desks, country missions, and country pro grams. Invariably there is a tendency—even though factors besides coun try size obviously affect aid allocations—for little countries to get large transfers per capita than big countries.

India is the case that proves this principle. From the very beginning of the post–World War II aid effort it has been the most consistently and seriously aided country. And yet in virtually any compilation of country aid receipts per capita during the past forty years India, because it is so much larger than other recipients, winds up at or near the bottom of the list.

IV. Summing Up

The present discussion has focused on the purely numerical—more particularly, the demographic—dimension of country size: What difference does it make to Indian development strategy and issues that the population is so large? Size also has another dimension: diversity. Exceptionally populous countries tend toward greater subregional (geographic, racial, ethnic, religious, linguistic) variety than other countries. Thus, all of the four 250-million-plus "giants" are, in fact, compound, multinational nation-states. It can be asked of each of them whether and how diversity complicates the maintenance of national cohesion. But that interesting question must be left for another time.

Meanwhile, as to the issue of scale, the Indian case does indeed support

the hypothesis that giant economies march to a significantly different drummer. Their management structures incline toward a cumbersomeness that is embedded in the system's extra hierarchical layering. This handicap invites two kinds of coping.

Done with adequate concern for the protection of national goals, downward delegations of politico-administrative decision can quicken management reflexes, promote popular participation, and encourage pluralism and policy experimentation in the system generally. Downward delegation faces an uphill struggle in Indian local governance, however, because of the country's tradition of decentralization by indoctrination.

Where leaders and officials are enlightened enough to reduce direct controls and their attendant rents, greater reliance on the ordering of economic decisions by well-framed markets (as well as more scope for other servomechanisms) can enhance growth and productive efficiency. In giant systems efficiency is driven more by internal competition, and ratios of external trade to real income and output are inherently low. But this is no excuse for uncritical turning inward—the easiest trap for the giant to fall into, especially in an era in which it must export something extra to pay for the extra imports of technology-embodying goods it needs.

Its large numbers afford the giant a large pool of talent from which to draw a disproportionately small number of top leaders. Large numbers facilitate the training of elites and may inject the positive effects of critical mass into the development of professional cadres.

In the realm of multilateral governance and official transfers, a giant economy such as India is victimized by the "state is a state is a state" syndrome. But we should not weep too much for the Indians. If they did not, on balance, prefer their giantism, there would be a general movement in the country, not just to break off corners like Kashmir, but to dissolve the compound state into a dozen Thailands.

ROOM TO MANEUVER IN THE MIDDLE

Local Elites, Political Action, and the State in India

by SUBRATA KUMAR MITRA*

INDIA'S remarkable political stability and high degree of participation have made her the odd one out among postcolonial states. The record of achievement is rather mixed in economic development, where slow but incremental growth is under considerable strain to keep pace with the growing expectations of a politically aware citizenry.[1] Yet even there, notwithstanding caste conflict and collective protest—*dharnas, boycotts, rasta roko*[2] and other forms of popular struggle that take after the Gandhian *satyagraha* of the 1920s—the basic stability of the state remains largely intact.[3] The legacy of public services, entrenched institutions before independence, and continuity of leadership have no doubt contributed to the stability of the postcolonial state.[4] Instead of dwelling

* A personal research grant from the Nuffield Foundation and comments from David Booth, Bruce Graham, Jack Hayward and four anonymous referees on an earlier draft of this paper are gratefully acknowledged.

[1] See Francine Frankel, *India's Political Economy, 1947–1977: The Gradual Revolution* (Delhi: Oxford University Press, 1978); Francine Frankel and M. S. A. Rao, *Dominance and State Power in India: Decline of a Social Order* (Delhi: Oxford University Press, 1989); and, for the political implications of poverty and redistribution in India, see Atul Kohli, *The State and Poverty in India* (Cambridge: Cambridge University Press, 1985).

[2] For a discussion of the scale and intensity of *gherao* (literally, to surround a decision maker) and its progeny—*rasta roko* (obstructing roads to disrupt traffic), *morcha* (militant procession), mass casual leave, *dharna* (refusal to clear an area when ordered to do so), etc.—see Robert Hardgrave and Stanley Kochanek, *India: Government and Politics in a Developing Nation* (New York: Harcourt, Brace, Jovanovich, 1986), 156; James Manor, "Collective Conflict in India," *Conflict Studies* 212 (London: The Centre for Security and Conflict Studies, 1988), 4. Geoffrey Moorhouse's *Calcutta: The City Revealed* (Harmondsworth: Penguin, 1974), 213–14, captures the form and spirit of *gherao* and its other variations, which are part of a new, authentically Indian genre of collective protest.

[3] With reference to secessionist violence in Punjab, three points should be noted. (1) The issue of legitimacy of the state in India continues to divide Sikh opinion; (2) though contested by terrorism, political participation continues to be perceived as the basis of legitimacy; and (3) the state still holds the upper hand in choosing its strategy in transactions with political actors. For the resilience of democracy in India, see Atul Kohli, ed., *India's Democracy: An Analysis of Changing State-Society Relations* (Princeton: Princeton University Press, 1988); and Myron Weiner, *The Indian Paradox: Essays in Indian Politics*, ed. Ashutosh Varshney (Delhi: Sage, 1989), who singles out the party as the key factor (p. 330).

[4] For a discussion of state formation in India, see Subrata Mitra, "Between Transaction and Transcendence: The State and the Institutionalisation of Power in India," in Subrata

on these factors I shall explore the universe of India's local elites for its contribution to the growth and legitimation of the authority of the state. This article suggests that protest and participation are but two facets of the complex repertoire of political man.[5] Although conventional political theory has treated them as distinct and contradictory forms of action, the view proposed here suggests that, under certain conditions, collective protest and the normal channels of political participation may be complementary forms of action. Under those circumstances, a state linked to society through multiple channels of articulation can expect greater legitimacy than can others that, in a similar situation, treat participation and protest as strictly separate and that favor "normal" forms of participation to the exclusion of all types of protest having an element of coercion in them. In order to examine this general proposition, this article analyzes the perception of India's local elites to the range of alternatives open to them in their dealings with the developmental bureaucracy.[6]

Local elites (or *gaon ka neta* as they are generically called in rural India) are political actors, situated at the interface of the modern state and traditional society. They are engaged in a two-way interpretation of norms and political demands. Their methods include acting as brokers in national, regional, and local elections; lobbying and contacting the bureaucracy and higher political elites; and collective protest. Their "room to maneuver" is created when bureaucrats, political decision makers, and the law of the land accept the legitimacy of multiple modes of participation from below. This, in turn, contributes to the expansion of the social base of the state through the involvement of a broader range of citizens in the process of social change. The ability of local elites to draw newly emerging social forces into the political arena and the policy

---

Mitra, ed., *The Post-Colonial State in Asia: Dialectics of Politics and Culture* (Hemel Hempstead: Wheatsheaf, 1990).

[5] The findings reported in Joel D. Barkan and Frank Holmquist, "Peasant-State Relations and the Social Base of Self-help in Kenya," *World Politics* 41 (April 1989), 359–80, provide cross-cultural support for this conjecture. They present the "marriage of peasant and state in self-help" as a dynamic factor that has "imbued that system with a measure of legitimacy, and forced the Kenyan state to be minimally accountable to the public . . ." (pp. 361, 363).

[6] In a different context, Sidney Tarrow has drawn our attention to the political and social function of local elites: "Far from behaving like traditional local notables or acting merely as rubber-stamp administrators of programs initiated from above, local elites have become important actors in the political adaptation of their communities to social and economic change." However, the larger implications of their social and economic roles and the methods they employ to achieve their objectives have not been widely noticed, particularly in the context of developing countries. See Tarrow, *Between Center and Periphery: Grassroots Politicians in Italy and France* (New Haven and London: Yale University Press, 1977). Some studies of *panchayatiraj* in India have raised this problem at the empirical level. See Iqbal Narain et al., *Panchayati Raj Administration* (New Delhi: Indian Institute of Public Administration, 1970).

process, and to exclude the traditional notables, has cast them as a conduit for the rapidly expanding postcolonial state.

Order, Development, and the Polarization Thesis

Some of these propositions are at variance with the earlier literature on conflict, protest, stability, and the strategies of development. Following the publication of Huntington's influential *Political Order in Changing Societies* (1968), the relationship between collective protest and legitimacy of state power has been considered to be generally negative.[7] On the basis of the findings of Hoselitz and Weiner,[8] Huntington suggested that rapid economic development in societies with inadequate institutional infrastructure can lead to violence, unrest, and loss of legitimacy. This theoretical linkage has been further reinforced by Gurr's notion of relative deprivation, which suggests precisely why discontent is felt most intensely in areas and among people experiencing rapid change.[9] These conjectures suggest a series of linkages among economic development, protest, and loss of legitimacy that can be specified in the form of a "polarization" model (Figure 1).

The assumptions and implications of the polarization model[10] have

		(1)		(2)		(3)
Economic Development	⟶	Inequality; Relative deprivation	⟶	Political conflict	⟶	Weakening of legitimacy

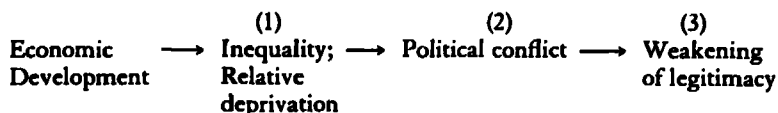



Figure 1
The "Polarization" Model

[7] This is the argument that runs through Samuel P. Huntington, *Political Order in Changing Societies* (New Haven: Yale University Press, 1968). It has influenced a number of works on this problem. On antisystem party strength and cabinet instability, see Michael Taylor and V. M. Herman, "Party System and Party Government," *American Political Science Review* 65 (March 1971), 28–37. On Communist Party membership and political instability, see Douglas Hibbs, *Mass Political Violence* (New York: John Wiley, 1973); and, on strength of extremist parties and reduced cabinet durability, see Lawrence Dodd, *Coalitions in Parliamentary Government* (Princeton: Princeton University Press, 1976). In his subsequent work Huntington has increasingly tended to see collective protest as a factor that can lead to reform and institutional realignment. See Samuel P. Huntington, *American Politics: The Promise of Disharmony* (Cambridge: Harvard University Press, 1981), esp. 203–10.

[8] Bert F. Hoselitz and Myron Weiner, "Economic Development and Political Stability in India," *Dissent* 8 (Spring 1961), 173. Huntington specified his model in terms of a series of causal statements that link social mobilization, discontent, and political instability; Huntington (fn. 7, 1968), 55.

[9] See Ted Robert Gurr, *Why Men Rebel* (Princeton: Princeton University Press, 1970) for the empirical specification of this concept.

[10] The polarization model is very popular with India's indigenous Left and their Western

been brought into question by Lichbach, who reports, on the basis of a comprehensive review of the literature, that no conclusive evidence of a relationship between economic inequality and political conflict exists.[11] Recent historical research has also challenged many previously well established assumptions about the reaction that developmental change was supposed to invoke in peasant societies. Rather than the spontaneous outburst of righteous—even mindless—anger, one is more likely to witness strategic protest from rational peasants, who try to combine both conventional and radical tactics to achieve their objectives. As Barkan and Holmquist suggest in their study of Kenya, "No longer regarded as passive actors in the state-peasant equation, peasants are now seen as rational actors seeking to maximize their individual, class and community interests within the multiple constraints of the social structure, the market, ecology and state policy."[12]

A Theory of Rational Protest

The general question that arises here is, what are the implications for the relations between protest behavior and state legitimacy once the state introduces, as intervening factors, policies of economic growth and redistribution in response to pressure from below? Is it not plausible to argue that the presence of an extremist party or group speaking in the name of the deprived and poised Janus-like at the fringe of the system—with an antigovernment public face but a private face organized around conventional lines—actually reduces bigger threats to the legitimacy of the system? Would this not be particularly true in the context of a responsive state? These speculations are given greater validity by the findings of Powell, who suggests that "riots and protests are less likely to

---

sympathizers. See Terry Byres, "The New Technology, Class Formation and Class Action in the Indian Countryside," *Journal of Peasant Studies* 8 (July 1981), 405–54.

[11] Mark Irving Lichbach, "An Evaluation of 'Does Economic Inequality Breed Political Conflict?' Studies," *World Politics* 41 (July 1989), 431–70. Lichbach questions the assumption that under the pressure of righteous anger arising out of relative deprivation people give vent to their collective frustration in a spate of irrational violence. In a different context, Charles Tilly has questioned such knee-jerk theories of political conflict. "The image," as Tilly remarks, commenting on these theories, "is hydraulic: hardship increases, pressure builds up, the vessel bursts. The angry individual acts as a reservoir of resentment, a conduit of tension, a boiler of fury. But not as a thinking, political man operating on principle"; Tilly, *The Formation of National States in Western Europe* (Princeton: Princeton University Press, 1975), 390.

[12] Barkan and Holmquist (fn. 5), 363. For applications of the same approach, see Robert Bates, *Markets and States in Tropical Africa* (Berkeley: University of California Press, 1981); and Samuel Popkin, *The Rational Peasant* (Berkeley: University of California Press, 1979).

appear under these conditions, perhaps because of the opportunity fc protest leaders to work within the system."[13] These conjectures can b put together as an alternative "rational protest" model (Figure 2).

	(1)	(2)	(3)	(4)
Economic development →	(a) Inequality (b) Relative deprivation →	(c) Contacting (d) Collective protest →	State responsiveness (Redistributive policies) →	Legitimac

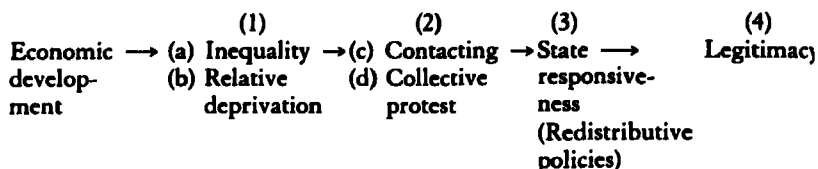



FIGURE 2
A "RATIONAL PROTEST" MODEL OF PARTICIPATION

In the first instance, the relative deprivation caused by unequal distri bution of the benefits of development intensifies pressure on the existin institutions and decision makers to accommodate the demands arisin from the newly mobilized social groups (Figure 2-1). This pressure ca take the form of various types of contact with authorities, lobbying, an protest action (Figure 2-2). One of the critical determinants of the specifi combination of these tactics is state responsiveness (Figure 2-3); the mor unyielding the state, the more radical the collective protest gets. Throug the accommodation of some strategic demands, however, the state ca enhance its legitimacy and take the wind out of the sail of revolutio (Figure 2-4); bring pressure to bear on its lower-level functionaries to remain responsive (Figure 2-3); initiate redistributive policies; and reduce *perceived inequality*.

The dual role of local elites, critically important to the process of rational protest, is, however, little understood in cross-cultural studies of development. By a curious coincidence, the concept that there is "room to maneuver" is anathema both to the classic development planners as well as to advocates of peasant revolution. Local elites are the bêtes noires of the advocates of the liberal-modernization approach[14] as well as of the radical-structural critics of that approach.[15] The former, top-down theorists of development, have routinely assumed that knowledge, technology, and capital accumulating at the top of the system will trickle down to the bottom in due course through the process of modernization. The

[13] G. Bingham Powell, "Extremist Parties and Political Turmoil: Two Puzzles," *American Journal of Political Science* 30 (May 1986), 374.

[14] R. Reddy and G. Hargopal, "The 'Pyraveekar': The Fixer in Rural India," *Asian Survey* 25, no. 1 (1985), 1148–62; G. Elwert and T. Bierschenk, "Development Aid as an Intervention in Dynamic Systems," *Sociologia Ruralis* 28, nos. 2 & 3 (1988).

[15] See Frankel (fn. 1, 1978).

ilure of the nature and course of economic development to live up to
ese expectations is blamed on the meddlesome middlemen who ob-
ruct the functioning of rational plans. Conversely, the advocates of rad-
al reform see in local elites an ally of repressive structures who defend
ickward social and economic systems. The main thrust of development
eory avoids mentioning local elites altogether, except to point to them
the enemies of economic development and social change.

One exception to this generally bleak picture is the remarkably per-
ptive study of Esman and Uphoff.[16] On the basis of a cross-national
ialysis of Asian experience, they conclude that effective local organiza-
ons "accountable to the local people, and involved in rural development
nctions, generally accomplished rural objectives more successfully with
spect to the available resource base than those with less rural organi-
tion."[17] However, while their analysis draws attention to the potential
r significant intervention at the local level in the process of develop-
ent, the focus on formal bodies shifts attention away from the local
tes, who are the real actors behind local organizations.

The rational protest model has potential applicability outside devel-
ping societies as well. Indeed, even in Great Britain—that epitome of
able democracy—"protest behaviour, far from being the occasional
utbursts of a hopelessly alienated minority, is an integral part of political
onsciousness and is viewed, under a variety of circumstances, as a legit-
nate pathway of political redress by widely differing sections of the
ommunity."[18] Cross-cultural support for this view is provided by the
fective, rather than deviant, character of "unconventional political be-
aviour,"[19] "disorderly politics,"[20] "deliberate defiance of the law,"[21] and
e efficacy of protest movements as complementary forms of political
ction in France.[22] Further supportive evidence, which would radically
nallenge the dismissal of protest behavior as destructive of the legiti-
acy of the state, is provided by Powell, who reports extremist party

[16] Norman Uphoff and Milton Esman, *Local Organizations for Rural Development: Analysis f Asian Experience* (Ithaca, N.Y.: Rural Development Committee, Cornell University, 1974).
[17] Ibid., ix.
[18] A. Marsh, *Protest and Political Consciousness* (Beverley Hills: Sage, 1977), 39.
[19] M. Kaase, *Political Ideology, Dissatisfaction and Protest* (Mannheim: Institut für Sozial-issenschaften an der Universität Mannheim, 1972); and E. M. Muller, "A Test of Partial heory of Potential for Political Violence," *American Political Science Review* 66 (September 972), 928–56.
[20] D. von Eschen, J. Kirk, and M. Pinard, "The Conditions of Direct Action in Democratic ociety," *Western Political Quarterly* 22 (June 1969), 309–25.
[21] J. Bondurant, *The Conquest of Violence: The Gandhian Philosophy of Conflict* (Berkeley: niversity of California Press, 1965); A. Carter, *Direct Action and Liberal Democracy* (Lon-on: Routledge and Kegan Paul, 1973).
[22] Philip Cerny, ed., *Social Movements and Protest in France* (London: Frances Pinter, 1982).

support to be associated with both cabinet instability and citizen turmoil, but not with "deadly violence or the overthrow of democratic regimes."[23]

Situating the Model in the Indian Context

A brief account of the role and history of the Indian state is helpful in the empirical specification of the abstract model proposed here. In contrast to Europe, the state in India is neither the product of a rising bourgeoisie leading an industrial revolution nor the product of the revolutionary seizure of power by political forces committed to a socialist revolution. At the center of the postindependence state were politicians who, during the long years of colonial rule, had perfected the art of sharing increasing amounts of power through competition and collaboration with the British. Independence further expanded the scope of political bargaining with competing interests that had become the hallmark of the Congress Party. "Development," understood as a generic expression for a series of mostly government-sponsored activities—the construction of roads, bridges, public utilities, and, more generally, social welfare, subsidies, and rapid economic growth—become the leitmotif of political discourse of postindependence India. Political formations of the Left, Right, and the Center confronted one another in successive general elections with their specific interpretations of the "developmental imperative." Under growing political pressure, the state in India has increasingly intervened in the social process by creating new social legislation and providing safety nets—or at least promising to do so. Whereas in Western Europe during the formative phase of capitalism, the state behaved very much as the "executive committee of the bourgeoisie," the state in India, like the avatars of Vishnu, reserved for itself a multiplicity of roles: of accommodation, extraction, production, and repression, stepping in as the inevitable intermediary whenever the conflict between social forces became threatening to public order.[24] Thus, the intermediary role of the state offered room for maneuver to vulnerable groups.

Local political process in India is molded by the conflicting demands of a modern state and a resilient traditional society; of a growing capitalist market set within a legal structure that is formally committed to egalitarianism. Since the general statements made here hang crucially

[23] Powell (fn. 13), 357.

[24] The powerful and numerous group of "bullock capitalists"—"small-scale, self-employed farmers who rely more on family labor and their own capital than on wage workers or machines"—is indicative of both the vitality of the process of capital formation and the effectiveness of social limits to the growth of predatory capitalism. For the multiple roles of the state in India, see Lloyd Rudolph and Susanne Rudolph, *In Pursuit of Lakshmi: The Political Economy of the Indian State* (Chicago: University of Chicago Press, 1987), 400–401.

upon the local elite, devising a field strategy for their identification and the measurement of their attitudes is of utmost importance in the empirical specification of our abstract model. Previous Indian research is of limited utility here. Based often on studies of single villages, it provides insight but is not helpful to answer theoretical questions or to draw general inferences. Studies of *panchayatiraj* (village councils) that utilize survey methods are more promising, but are not very helpful in terms of illuminating the political world of local leaders because, more often than not, they are based only on interviews with the local officials.

I therefore decided at the beginning of my fieldwork to utilize the reputational technique to identify local elites.[25] Thus, in each of the eighteen villages studied, five informants were asked to list the *gaon ka netas* in their order of influence. These lists, compiled independently, were then compared, and the names appearing most frequently were retained as a working roster of local elites in a particular village. In this process, 233 people were identified from among the ranks of those within the *panchayatiraj* structure (village and area councils), as well as from cooperatives, school boards, school teachers, and Village Level Workers (who, as government employees, are not entitled to dabble in local politics but do so all the same), and caste organizations. These leaders were interviewed on the basis of both structured and open-ended questions on a range of topics that covered the problems faced by the people of the area, the solutions they considered desirable, the institution they considered best placed to implement these solutions, and the variety of methods they pursued to influence bureaucrats and political leaders at higher levels. The interviews, conducted during 1979–80 simultaneously with a larger survey of a cross-section of the population in the study villages, were undertaken by the author with the help of research assistants and were conducted in the local languages.[26]

[25] For the application of the reputational technique for elite studies in India, see Rajendra Singh, *Land, Power and People: Rural Elites in Transition, 1808–1970* (Delhi: Sage, 1988).

[26] Local protest movements came into prominence in the early 1970s following the deinstitutionalization of the Congress Party and the decline of intermediate structures. The fieldwork for this study was conducted after the lifting of the national Emergency (1975–77), which had imposed severe restrictions on the freedom of political activity. The full range of political activities had been restored for about two years when the survey was conducted. The combination of plebiscitary politics at the national level and protest movements at the regional and local level that characterized the spatiotemporal context of the survey have continued to be the dominant themes of Indian politics through the 1980s. (See Rudolph and Rudolph [fn. 24], 127–58, for the rise of plebiscitary politics, and 247–58 on demand groups.) For the results of the survey of the cross-section conducted simultaneously with the elite study, see H. R. Chaturvedi and Subrata Mitra, *Citizen Participation in Rural Development* (Delhi: Oxford and IBH, 1982). A comprehensive report on the elite data, which continue to be interesting because of the enduring importance of local protest movements and the distinctive styles of Gujarat and Orissa politics, will be published in Subrata Mitra, *Power, Protest and Participation* (London: Unwin, Hyman, forthcoming).

Any study purporting to suggest a generalization about India must come to grips with India's regional diversity. State responsiveness and extensive political bargaining are more likely to be present in those parts of India that are more advanced in terms of economic and infrastructural development. A second contributory factor to this diversity is the length and depth of British colonial rule, for the "due process of law" and limited political participation that the British made available to their subjects, even under colonial conditions, was not applicable to the natives of territories under the rule of Indian princes. In the empirical specification of the model, therefore, the regional variation is incorporated through the selection of two regions of contrasting character: Gujarat, with a long history of participation and economic development, and relatively backward Orissa, large parts of which were ruled by Indian princes and were integrated with mainstream politics and the market only after independence.

The samples for the survey were chosen from Surat and Dhenkanal, two districts at the median level of development, respectively in Gujarat and Orissa. Using a quasi-experimental research design that controlled for village size, nine villages with high, medium, and low levels of development were selected in each district. Surat district in South Gujarat has a long history of capitalist development. The rural scene has been virtually transformed from a subsistence agriculture to one in which capitalist farming provides an important, but not the primary, source of occupation to the majority of the rural population. Dhenkanal, on the other hand, provides a contrasting rural scene, where sporadic attempts to introduce new agricultural technology have created isolated islands of modern farming surrounded by a vast hinterland of subsistence agriculture. There is nevertheless a sense of movement in the labor market, owing to the policy of rapid growth of rural industries undertaken by the Orissa state government. These activities have been drawing away some agricultural workers, creating resentment among employers and strains within the traditional social structure where a *halia* (farm servant) used to work for a given master on the basis of annual contract. As such, the situation is one of radicalized workers, often pitted against employers operating within a relatively traditional economy.[27] Despite changes in the high politics of Delhi, recent studies have confirmed the enduring nature of the contrast between the two regions.

[27] See India, *Techno-Economic Survey of Orissa* (New Delhi: National Council of Applied Economic Research, 1962); and idem, *The Techno-Economic Survey of Gujarat* (New Delhi: National Council of Applied Economic Research, 1963), for the socioeconomic background of Orissa and Gujarat.

other residents of the village, local elites are affected by the na- scale of development in their village. They attempt to overcome ate allocation to the village through negotiation with decision at higher levels. Alternatively, and sometimes in addition, they essure through popular mobilization, petitions, and, eventually, iption of public order. In the analysis that follows, the data are l by region and, within each region, cross-tabulated by social cial stratification is measured in terms of caste. The so-called :es, especially Brahmins, traders, and writer castes who, during ule, capitalized on their links with the outside world and entered :ssions and government services, are classified as upper caste. For l analysis, the cultivating castes are grouped together as the mid- s, and the former untouchables and indigenous population (tri- counted as Adivasi/Harijan.

the perspective of the polarization model, economic develop- ects the three social classes quite differently. Those with contacts quate capital (in this case, the upper and the middle castes) can ted to do rather nicely from development, whereas the Adivasi/ are vulnerable to a rapid erosion of their standard of living as ng space is "enclosed" by a rapidly advancing capitalist economy ıe sources of supplementary income are steadily removed by the ional bookkeeping of the modern economy.[28] The polarization vould lead to the expectation of greater relative deprivation he Adivasi/Harijan than among the upper castes, but the expec- f the rational protest model would be more complex and mixed cter. Further, the expectation arising out of the polarization 'ould be one of intense relative deprivation, conflict, and loss of :y, whereas the rational protest model would channel relative ion towards political action. The protest model would witness : use of "contacting" and all forms of social and political lever- ıding various types of collective protest that, although coercive, less stop short of outright revolution. Some of these hypothetical ons are examined below in light of the empirical findings.

Some Partial Evidence of the Room to Maneuver

n to maneuver in the middle as defined above is not a directly ɔle variable, but its presence can be observed in terms of its im-

of course, be argued here that the growth of labor markets should offset some of p faced by those without any capital except their labor. However, the migration eople can in effect send both the prevailing wage rate and the annual availability

plications. The following analysis will examine two sources of evidenc the social origin of local elites and their own evaluation of the modes c political action.

The Social Profile of the Local Elite

The sample of local elites manifested a fair amount of social and eco nomic diversity, with sizable representation of Scheduled Castes, Sched uled Tribes (so-called, because they have been named in the ninth Sched ule of the Constitution of India), as well as of marginal and smal farmers. (See Table 1.) The results presented in Table 1 throw into ques tion some of the arguments of the polarization model, which asserts tha leadership at the local level is monopolized by the "dominant elites" wh possess land, money, power, and high social status. As such, they manag to corner most of the benefits of development. The field data, in reveal ing considerable social heterogeneity in the profile of local elites, questio the rather simplistic view of the polarization model, which takes int consideration neither the diversity in social origin of local elites nor th fragile and variable base of their political support, contingent in mos cases on their ability to promote the interests of their clientele.

Of particular significance is the presence of large numbers of Adivasi Harijan among local elites. Growing politicization of the Indian elector ate has led to the steady broadening of the political arena, considerabl turnover of leadership, and the entry into the political process of thos who were hitherto excluded.[29] The results presented in Table 1 supple- ment this growing body of electoral evidence by revealing the social bas of local elites. While the age profiles of the two regions are roughly com- parable, the proportion of upper-caste people among local elites is less in more developed Gujarat than in rustic Orissa. Former untouchables and tribals have made greater headway in Gujarat. About a fifth of the local elites in both regions are illiterate or barely literate, and a significant number have less than five acres of land. These and the other social in- dicators reported in Table 1 reveal the extent of social heterogeneity

---

of work downwards. See Jan Breman, *Of Peasants, Migrants and Paupers: Labour Circulation and Capitalist Production in West India* (Delhi: Oxford University Press, 1984), xx.

[29] Supportive evidence of this form of mobility is provided by the literature on electoral mobilization, which informs us of the remarkable progress made in several Indian states by the "backward classes," which in the politically turbulent sixties, were successful in dislodg- ing the upper-caste leadership from their dominant position. Evidence of electoral mobili- zation of agrarian interests through "backward caste movements" is provided by Rudolph and Rudolph (fn. 24). Also see the contributions of John Wood and Rhoderick Church in John Wood, ed., *State Politics in Contemporary India: Crisis or Continuity?* (Boulder, Colo., and London: Westview Press, 1984), for detailed discussions of the state of mobilization of the "backward classes" in various Indian states.

TABLE 1
SOCIODEMOGRAPHIC PROFILE OF LOCAL ELITES IN GUJARAT AND ORISSA
(PERCENTAGE)

	Gujarat (N=131)	*Orissa (N=102)*
Age		
Under 34	16	18
35 to 49	41	44
50 and above	43	38
Caste		
Upper	16	31
Middle	37	46
Adivasi/Harijan	42	22
Non-Hindu	5	1
Education		
Illiterate or barely literate	19	15
Primary or middle level	41	55
Matriculated or above	40	30
Occupation		
Owner/cultivator	65	69
Agricultural worker	9	4
Small trader or professional	22	20
Other	4	7
Land ownership		
Less than 5 acres	18	45
6 to 20 acres	49	43
More than 20 acres	33	12
Organizational involvement		
None	5	7
1 to 2 organizations	41	47
3 to 5 organizations	31	36
More than 5 organizations	23	10

among local elites, both in a relatively developed as well as in a backward region of India. Subsequent tables will reveal the complex interplay of these factors as they affect the flow of material resources and attitudes toward participation.

MODES OF POLITICAL ACTION

What kinds of action do local leaders approve of as means of putting pressure on decision makers? From Table 2, it appears that the leaders in both regions overwhelmingly approve of such conventional forms of

action as contacting important leaders and officials or making representations to higher authorities. The pattern, when it comes to radical action, is much more complex. "Hard" forms of direct action such as nonpayment of taxes and *gherao* (surrounding a decision maker and holding him until a favorable decision is obtained), which have larger potential costs in terms of individual penalty, carry less support than "softer" forms such as *morcha* (a procession intended as a show of strength). Only a minority appears to approve of radical collective action in most cases.

TABLE 2
NORMATIVE APPROVAL OF MODES OF PARTICIPATION,
CROSS-TABULATED WITH CASTE
(PERCENTAGE)

Question: *It has been seen that people engage in different kinds of acts for drawing attention of concerned authorities to their demands and grievances. Please tell me which of these acts you personally approve of or disapprove of.*

Respondent approves of	*Gujarat*	*Orissa*
Contacting important leaders	97	98
Contacting important officials	95	98
Representations like signature campaigns or memoranda	95	94
Joining demonstrations	21[a]	40
Morcha (a gathering intended as a show of strength)	39	31
Strike	13[b]	16
Nonpayment of taxes	11	8[c]
Gherao (surrounding decision makers to put pressure)	7[d]	12

[a] Correlation with caste status (Kendall's tau b) = 0.12, significant at 7%.
[b] Correlation with caste status (Kendall's tau b) = 0.22, significant at 0.04%.
[c] Correlation with caste status (Kendall's tau b) = 0.16, significant at 4%.
[d] Correlation with caste status (Kendall's tau b) = 0.14, significant at 5%.

How effective is radical action perceived to be as a method of material advancement? The responses to this question are presented in Table 3. In both states, people from the upper social stratum are less likely to approve of protest and agitation. From the cross-tabulation of the data presented in Table 3 with social class, the Adivasi-Harijan both in Gujarat and Orissa—at 57 percent and 50 percent, respectively—believe that conditions can be improved through protest and agitation. In both Gujarat and Orissa the upper stratum tends to believe that protest and agitation hinder development, but the pattern of their disapproval is different. While in Orissa both the strata are equally opposed to protest and agitation, at 79 percent, the hostility of the Gujarati middle stratum to "protests and agitations by the poor and the depressed"—12 percent ahead of the uppermost stratum—is reminiscent of Breman's reports of

TABLE 3
IMPACT OF PROTEST AND AGITATION ON DEVELOPMENT, CROSS-TABULATED WITH CASTE
(PERCENTAGE)

Question: *Some people say that it is through protest and agitation that the poor and depressed will improve their conditions. Others say that protests and agitations of this kind hinder development and will not benefit anyone. What do you think?*

	Adivasi/ Harijan	*Middle Caste*	*Upper Caste*	*Total*
Gujarat respondents ($N = 130$)[a]				
Improve conditions	57	21	33	39
Hinder development	43	79	67	61
Orissa respondents ($N = 102$)[b]				
Improve conditions	50	23	25	30
Hinder development	50	77	75	70

[a] Kendall's tau b = 0.23, significant at .03%.
[b] Kendall's tau b = 0.18, significant at 2%.

the intense antagonism of the landowning middle social stratum toward the lowest social groups.

SOCIOSTRUCTURAL AND POLITICAL FACTORS IN THE BENEFITS OF DEVELOPMENT

Having already examined some partial indicators of attitudes toward normal and radical modes of participation, we shall now offer a more systematic examination of this process. This is done by first formulating the implications of the polarization model in terms of a "structural" equation that suggests that the higher the social class, the greater are the gains one might expect out of development. The model is specified in terms of a simple equation:

Equation 1

Benefits of development = f (social class)
where $f > 0$,
social class is measured by caste status, size of landholding, and education.

Benefits of development are usually measured in India in terms of a variety of "objective" indicators pertaining to material resources such as agricultural inputs and developmental loans. We expanded this definition by including a more subjective "satisfaction" with development as

well as improvement in the overall financial situation. From the foregoing, we can conceptualize the routes one might traverse to reach these desired goals. The first of these requires the least amount of political or organizational effort. This route, "participation without politics," is secured through the normal operation of the market.[30] Ceteris paribus, those with the necessary initial advantages of wealth, social standing, and education are likely to be the normal beneficiaries of participation without politics as implied by Equation 1. There are, however, other routes to participation that are more political in character. These consist in contacting leaders and officials at higher levels, getting involved in organizations, and taking direct action. While they are conceptualized in the abstract as independent paths, in the real world of politics they are likely to operate in complex combinations. To measure their independent and joint contributions to participation in development, we can specify a structural-political model in terms of a second equation:

Equation 2

Benefits of development = f (social class, organization and contacting, attitudes to collective protest)
where $f > 0$,
- (a) social class is measured as above,
- (b) organization is measured in terms of the number of organizations with which the respondent is associated,
- (c) contacting is measured in terms of the frequency with which the individual contacts decision makers at higher levels, and,
- (d) attitude to collective protest is measured in terms of the local elites' approval of direct action.

"Benefits of development," the major dependent variable in this analysis, was measured in terms of a scale created by totaling responses to questions about items such as improved seeds, chemical fertilizer, irrigation, house sites, the government "land to the landless" program, bank loans, subsidies, improved agricultural implements, and government aid for promotion of small industries. For the purposes of constructing this scale of the benefits of development, those who had received any of the named items during the previous two years scored one point on each item; those who had not, scored zero. Considering the relative levels of development of Orissa and Gujarat, the results do not come as a surprise. Whereas in relatively backward Orissa only 14 percent of the local elites had received six or more of the items on the scale, the corresponding

[30] Samuel Brittan, *Participation without Politics: An Analysis of the Nature and Role of Markets* (London: Institute of Economic Affairs, 1975).

score for Gujarat was 30.5 percent. In both Gujarat and Orissa, benefits of development were closely related to satisfaction with improvement in the financial position of the respondent (with a correlation of .4 in Gujarat and .18 in Orissa) and in general satisfaction with development of the area. (For this, the correlation is .21 in Gujarat and .27 in Orissa.) The subjective indicators themselves were positively correlated, giving both measurements of development a reasonable degree of internal coherence.

Using the above parameters for both the structural and the structural-political models, Table 4 presents the results of the multiple regression in order to test Equation 1, whose conjecture it only partly confirms. Thus, the positive coefficient of caste status in Orissa in regression equation 4.1 (in Table 4) confirms the hypothetical expectation that the lower the caste, the less are the chances of getting developmental benefits.

TABLE 4
SOCIOECONOMIC FACTORS ASSOCIATED WITH THE BENEFITS OF DEVELOPMENT
(STANDARDIZED BETA COEFFICIENTS)[a]

Independent Variable	*Dependent Variable*			
	Material Benefits	*Present Financial Situation*	*Finances Compared to the Past*	*Overall Satisfaction with Development*
Regression Equation	4.1	4.2	4.3	4.4
Gujarat respondents (N = 131)				
Social class[b]	−.31	.2	−.33	—
Education	—	—	—	.23
Land	—	.24	—	.26
R^2	.11	.13	.09	.12
Orissa respondents (N = 102)				
Social class[b]	.25	—	.31	—
Education	—	—	—	−.31
Land	—	—	—	—
R^2	.06	.01	.10	.08

[a] Only coefficients that are significant at .05 level or less have been reported.
[b] Social class is measured in terms of caste status, scaled as follows: Adivasi/Harijan = 1, middle caste = 2, upper caste = 3.

Though the model explains only 6 percent of the variance, the statistica relation is in the predicted direction. High multicollinearity, caused b the strong correlation between land, caste, and education, reduces th significance of the coefficients of education and landholding. Howeve turning next to results from Gujarat in regression equation 4.1 (Table 4) it is immediately obvious that, rather than low caste status reducing one' chances of obtaining developmental benefits, the strongly negative coef ficients suggest that it actually improves them. The variance of 11 per cent is modest but higher than that explained by the same model in Orissa. Moving over to equation 4.2, we notice that the model does no produce any significant coefficients in Orissa. However, in Gujarat, th implications of the model are somewhat contradictory. Those ownin land are more likely to be satisfied with their present financial situation but the higher the caste, the opposite is more likely to be the case. Fro equation 4.3 in Orissa, we see that the lower the caste, the less one i satisfied with one's present financial situation; whereas the opposite i true in Gujarat. Finally, as regards overall satisfaction with development in Orissa the more educated are less satisfied. In Gujarat, on the other hand, the data conform to Equation 1: the higher the education and the size of landholding, the more one is satisfied with the pace and quantum of development.

What happens when people do not receive the benefits they believe they are entitled to? In order to account for the contrast between Gujarat and Orissa, we need to add some political and organizational variables, as explanatory factors, through Equation 2. We know from Table 3 that more Gujaratis believe in the efficacy of protests and agitation than do the local elites in Orissa. Moreover, belief in the efficacy of development in Gujarat is not simply normative. It seems to have a firm basis in the developmental experience of the Adivasi-Harijan in the two states as well. In terms of both the tangible benefits of development and their general financial situation, these historically underprivileged groups in Gujarat seem to have had a more positive experience compared with those in Orissa. An impressive group of 60 percent of the Adivasi-Harijan in Gujarat are in the highest category on the composite scale of benefits, compared to 9 percent of the uppermost social class and 16 percent of the middle social class. The corresponding figures from Orissa are, respectively: 9 percent for the Adivasi-Harijan, 11 percent for the middle, and 22 percent for the upper social classes. Even more spectacularly, 91 percent of the Adivasi-Harijan questioned reported an improvement in their financial situation over the past ten or fifteen years in Gujarat, whereas none among the Adivasi-Harijan in Orissa reported havin ex-

perienced an improvement in personal financial situation over a similar period.

Among other factors that further enhance involvement with—and benefits from—the process of development are contacts with leaders and officials at higher levels, involvement with a range of intermediate organizations, and psychological preparedness to engage in radical behavior. At the level of partial correlations, each of these factors is positively related to developmental benefits in Orissa (respectively, .15, .14 and .22). By contrast, in Gujarat, only organizational involvement is positively related to benefits; and radical attitudes are expressed by those who have not been successful in securing benefits. Turning to the regression results presented in Table 5, though the addition of these variables causes the variance explained to go up modestly, their contributions are not significant. In regression equation 5.3, whereas organizational involvement in Gujarat contributes positively to the improvement in an individual's financial situation compared with the past, the opposite is true in Orissa. One explanation for this is that there are a considerable number of local elites from the Adivasi-Harijan in Orissa who are involved in a range of organizations but who complain bitterly that their financial situation has definitely worsened. Radical sentiments are expressed by those from the upper social stratum in Gujarat who have not seen any improvement in their financial situation over the past. (See 5.4 in Table 5.) However, those who are satisfied are also likely to take radical action to achieve their ends. In other words, the political repertoire of developmental participation in Gujarat includes conventional as well as radical action, used in an optimal combination. In Orissa, on the other hand, anger and frustration are expressions of a sense of personal impotence, more likely to be the rebellion of the deprived than a tactical instrument in the hands of resourceful political entrepreneurs.

MULTIPLE MODES OF PARTICIPATION

The "room to maneuver in the middle" implies that radical and conventional behavior complement one another. However, when society is polarized into antagonistic camps, one would expect a negative correlation between attitudes toward conventional and radical modes of action. Elite perception of the nature of the state is crucial to this relationship: on the one hand, the image of the state as a manipulable source of bounty will encourage attitudes closer to a transactional model; on the other hand, a state that is perceived as distant and paternalistic would not lead to a sense of individual efficacy. In the latter case, subjects are likely to ac-

TABLE 5
CONTRIBUTION OF STRUCTURAL AND POLITICAL FACTORS TO THE BENEFITS OF DEVELOPMENT
(STANDARDIZED BETA COEFFICIENTS)[a]

Independent Variable	*Dependent Variable*			
	Material Benefits of Development	*Present Financial Situation*	*Finances Compared to the Past*	*Overall Satisfaction with Development*
Regression Equation	5.1	5.2	5.3	5.4
Gujarat respondents (N = 131)				
Social class[b]	−.28	—	−.33	−.33
Education	—	—	—	—
Land	—	.3	—	—
Contacting[c]	—	—	—	—
Organization[d]	—	—	.22	.22
Radical norm[e]	—	—	—	−.21
R^2	.13	.18	.18	.19
Orissa respondents (N = 102)				
Social class[b]	.24	—	.32	—
Education	−.28	—	—	−.32
Land	—	—	—	—
Contacting[c]	—	—	—	.2
Organization[d]	—	—	−.24	—
Radical norm[e]	.21	—	—	−.2
R^2	.14	.02	.14	.15

[a] Only coefficients that are significant at .05 level or less have been reported.
[b] Social class is measured in terms of caste status, scaled as follows: Adivasi/Harijan = 1, middle caste = 2, upper caste = 3.
[c] Contacting is measured by adding up the frequency with which local elites contact leaders and officials at higher levels: never = 1, rarely = 2, sometimes = 3, frequently = 4.
[d] Organization is measured by adding up the number of organizations and associations with which the respondent claims to be involved.
[e] Radical norm is measured by adding up the approval modes of protest behavior: *morcha, gherao*, strike, nonpayment of taxes. Approval = 1, nonapproval = 0.

quiesce more frequently. However, when they are pushed beyond the limit, the outcome would be that of no-holds-barred rebelliousness.

The questions that were asked to measure the perception and evaluation by local influentials of the range of activities related to participation has already been discussed in Table 2. The relationship between conven-

tional and radical modes of action was tested on the basis of a factor analysis of modes of participation. The factor analysis was carried out on the sample as a whole, as well as on the local elites of Gujarat and Orissa separately. The results of the analysis on the full sample (Table 6) point to the existence of two principal factors, explaining 46 percent of the variance. On the basis of the factor loadings, we can identify two principal factors: factor 1 represents the conventional dimension of participation, consisting of "personal approval of signature campaigns" and contacting important leaders and officials; factor 2, with high loadings on various forms of coercive and direct action, constitutes the radical dimension.

The factor correlation between the conventional and the radical dimensions of participation is a relatively weak but positive .06. However, when carried out at the regional level, the model reveals interesting local variations. In Gujarat, the factor correlation between the conventional and the radical dimensions is positive, and, at .2, is stronger than that of the correlation on the total sample. In Orissa, the factor structure is slightly different. The radical dimension is composed of the same items as in Gujarat. The conventional dimension appears as two distinct but positively related factors, with signature campaigns appearing as a separate dimension. The correlation between the first dimension, composed of contacting leaders and officials, and radical dimension is, at −.26, quite strongly negative.

TABLE 6
FACTOR ANALYSIS OF THE MODES OF PARTICIPATION (N = 233)
(WITH OBLIQUE ROTATION)

Respondent Personally Approves of:	*Factor Loadings* *Factor 1*	*Factor 2*
Representations such as signature campaigns or memoranda	[.36]	.11
Joining demonstrations	.18	[.65]
Morcha, processions	.11	[.43]
Gherao	.16	[.90]
Strike	.04	[.77]
Nonpayment of taxes	.14	[.57]
Contacting important leaders	[.99]	.06
Contacting important officials	[.83]	.21

CONCLUSION

The Indian case has wider implications because it represents a significant attempt to achieve economic transformation through constitutional

means and the "rule of law."[31] The empirical problems discussed in th article point toward larger theoretical issues. What cross-cultural infe ences might one draw from the Indian experience? Does radical protes emerging out of a sense of injustice, diminish or extend the legitima of the state in India?

Local elites are the crucial piece in the puzzle that Indian politic economy presents to those who wish to make a short-term prognosi Conceptually, local elites operate as a "hinge" at the meeting point of th modern state and the traditional society. Their social composition an political character are critical to the shape of state-society relations.

Several scenarios are possible. Competing local elites, operating withi institutions that have at least a modicum of political autonomy and a representative and accountable to a larger constituency, can act as th cutting edge of an expanding, liberal state. However, local elites, presen as a socially cohesive group able to cumulate the advantages arising ou of the control of means of production, social status, and physical forc can use local institutions to reinforce their political dominance also. I terms of her empirical understanding of rural India, Frankel appears t be closer to the second view. Hence the skepticism about the ability o the state to extend its norms of social equality and political participatio into the domain of the traditional society. "[For the purpose of economi development and social change] institutional changes, in particula would have to be implemented in the villages through the existing loca authorities, that is, the very rural elites against whose power they wer directed. At the villages, the loudest voices still belong to members of th dominant landowning classes."[32]

In contrast to this image of a socially cohesive elite, dominating rura life through its control of the means of production and political powe our field data indicate that local elites are far from being socially homo geneous and politically cohesive. Allocation of developmental benefits i the local political arena is not the result of social dominance but is th outcome of a complex process based on bargaining and political comp tition, however imperfect. Thanks to the state structure and the politic process that characterizes it, discontent and relative deprivation arisin out of unequal allocation of developmental benefits do not result in ap athy and anomic powerlessness. Instead, people engage in various form of direct political action. All forms of intervention in the process of d velopmental allocation—ranging from contacting decision makers

[31] John Lewis, *Quiet Crisis in India: Economic Development and American Policy* (Gar City, N.Y.: Anchor Books, 1964), 8.

[32] Frankel (fn. 1, 1978), 547.

higher levels to direct action of indeterminate legality such as *gherao*—are seen as shades of legitimate political action. All these forms of action, related through normative approval and participation, constitute a complex repertoire. The strategic choice of a particular form of action is contingent upon the local configuration of power and issues at stake. Their actual and potential availability within the political system extends the set of relevant players and the developmental agenda, as well as the political arena within which the issues of allocation are settled.

Political trends in various states of India provide variations around this basic model. Both in Gujarat and in Orissa, the lowest social strata are very much present within the local political arena. But in Gujarat, the former untouchables and tribal leaders are part of a wider network, whereas similar organizations and political linkages are not available in Orissa. There, the absence of these institutional and political links and the slower rates of growth in the rural economy have combined to retard the pace of accommodating lower-class interest. By contrast, in Gujarat, the consequences of a rapidly expanding rural economy, combined with radical and organized lower-class politics, can be seen in the relatively more successful accommodation of the political agenda of the lower strata.[33]

Some scholars who view culture as a limiting factor for political concepts and action, pointing to the failure of constitutional experiments in the vast majority of postcolonial societies, might argue that cross-cultural export of good practices is essentially a neo-modernist pipe dream.[34] But the Indian case does offer a moral for those seeking to achieve, *simultaneously*, both stability of the state and some measure of economic growth within a democratic political structure. The Indian case demonstrates that room for vulnerable social groups to maneuver—room that can be created through decentralization of power, active political participation at the local level, and legal and administrative tolerance for protest behavior—can contribute to the legitimacy of the state.

Orthodox defenders of political stability for whom any protest is anathema might read at least a mild form of "Huntington's revenge"[35]

[33] On the mobilization of lower castes in Gujarat, see John Wood, "Congress Restored? The 'KHAM' Strategy and Congress (I) Recruitment in Gujarat," in John Wood (fn. 29); and, for upper-caste backlash against the policy of reservation in favor of the former untouchables, tribals, and backward classes, see Subrata Mitra, "The Perils of Promoting Equality: The Latent Significance of the Anti-Reservation Movement in India," *Journal of Commonwealth and Comparative Politics* 25, no. 3(1987), 292–312.

[34] See, for example, Victor Funnell, "The Philippines: In Search of the Nebulous State," in Subrata Kumar Mitra, ed., *The Post-Colonial State in Asia: Dialectics of Politics and Culture* (Hemel Hempstead: Wheatsheaf, 1990), 131–54.

[35] The reference here is to the trade-off between political stability and economic growth

into India's undistinguished record of economic growth compared to th economic miracle of the newly industrialized countries (NICs) of South east Asia. By accelerating the entry of new social forces into the politica arena, protest behavior can overstretch governmental capacity, beside cluttering the political agenda with demands with which the system i unable to cope. All of these contribute to the reduction of political coher ence in the system, leading to a political crisis.[36] The rapid mobilizatio of the intermediate castes in the 1960s caused similar crises in many o India's regional governments. However, this experience of the efficacy o protest might have provided the confidence and organization with whicl the opponents of Mrs. Gandhi's "Emergency" in 1975 united in thei opposition to authoritarian rule. In the last analysis, an important expla nation for the resilience of Indian democracy is found in the comple political repertoire of local elites and in their remarkable blend of con ventional modes of participation and collective protest.

Two caveats need to be specified here regarding the cross-cultural generalizability of these findings. In the first place, as we learn from the regional comparison undertaken in this study, the "Indian model" is not equally applicable all over India; for the policies of the regional government concerned, as well as its specific tradition of participation and administrative responsiveness, are crucial intervening variables. While the local elites of Gujarat have mastered the art of balancing negotiation and bargaining with collective protest, their counterparts in Orissa still perceive these as separate strategies. However, within India's federal structure, there is a tremendous amount of knowledge-sharing at the regional level. The eventual diffusion of the Gujarat model to other regions cannot therefore be ruled out.

Secondly, the high responsiveness of the Indian state, itself the outcome of a particular history of state formation, is a limiting factor on the cross-cultural validity of "room to maneuver" as a panacea for political legitimacy. The accommodative nature of the state in India facilitates the spread of transactional values that constrain the behavior of all actors engaged in the game of political bargaining. This specific attribute of the state in India originated with the success of the colonial state in linking its authority to the "elite entrepreneurial network" that characterized Indian society during the high noon of the empire.[37] The Congress Party,

that Huntington (fn. 7, 1968) described as options for developing countries. Thus, those developing countries that opt for mass participation and populist economic policies must pay the penalty of high inflation, economic stagnation, and heavy foreign debts.

[36] Powell (fn. 13), 374.

[37] Richard Sisson and Stanley Wolpert, *Congress and Indian Nationalism: The Pre-Independence Phase* (Berkeley: University of California Press, 1988).

product of the pas de deux between the Empire and the Indian middle classes, was eventually linked to the peasant base of the society under the Gandhian noncooperation movement, an enduring legacy that helped institutionalize participation as a cardinal value in modern Indian politics.

Within the limitations stated above, results presented here throw into doubt the general validity of the polarization thesis and of early modernization theories that assert that protest behavior necessarily leads to the decline of legitimacy of the state. It is more appropriate to think in terms of a critical but variable threshold of tolerance for protest behavior as a helpful factor for political legitimacy. To this end, enough power has to be conceded to local elites so that they may develop a stake in the institutions of the state. However, too much local power at the expense of the political center can obstruct the induction of new elites from previously powerless groups, causing, in the process, the transfer of authority from the state and the political arena altogether to those who have little patience or need for democratic niceties.

Review Articles

OLIGARCHS AND CRONIES IN T PHILIPPINE STATE

The Politics of Patrimonial Plunde

By PAUL D. HUTCHCROFT*

Belinda A. Aquino, *Politics of Plunder: The Philippines under Marc* zon City: Great Books Trading and the University of the Ph College of Public Administration, 1987, 208 pp.

Robin Broad, *Unequal Alliance: The World Bank, the International l Fund, and the Philippines.* Berkeley and Los Angeles: University fornia Press, 1988, 352 pp.

Gary Hawes, *The Philippine State and the Marcos Regime: The P Export.* Ithaca, N.Y.: Cornell University Press, 1987, 196 pp.

David Wurfel, *Filipino Politics: Development and Decay.* Ithaca, N. nell University Press, 1988, 361 pp.

Introduction

THE rise and fall of Ferdinand E. Marcos in the Philippin lights the important distinction that needs to be made l changes in political regime and changes in the nature of the state. nearly a decade and a half of authoritarian rule (1972–86), and years devoted to the restoration of pre–martial law institutions, of the state in the postwar Philippine political economy has perh played more fundamental continuity than discontinuity: the mor change, the more they stay the same. Access to the state appar mains the major avenue to private accumulation, and the quest fc

* The author wishes to acknowledge the assistance of the University of the F School of Economics, where he was based while conducting the dissertation res inspired this article. He was supported by a grant from the U.S. Department of Fulbright-Hays Doctoral Dissertation Research Abroad Program, as well as by a g the Social Science Research Council and the American Council of Learned Soci funds provided by the William and Flora Hewlett Foundation.

In addition, he would like to thank all those who offered suggestions and comi contributed to the ideas in this review article: Ben Anderson, Gerry Burns, Nick Edna Labra Hutchcroft, Alex Magno, Manuel Montes, Jim Rush, Jim Scott, and J as well as those who commented on an earlier version, read at the Third Internati ippine Studies Conference, Quezon City, July 13–17, 1989. The author, of cour responsibility for any errors or omissions within the article.

ıeeking" opportunities continues to bring a stampede of favored elites ınd would-be favored elites to the gates of the presidential palace.[1] A key]uestion, put very crudely by one observer, concerns "why the Philip-)ines could not use the dictatorship as its springboard for economic de- /elopment, as we witness in newly industrialized countries in Asia."[2] Alternatively, we need to ask why, despite two major regime changes in he postwar era (Marcos's declaration of martial law in 1972 and his over- hrow in 1986), there has been so little change in the way in which dom- nant economic interests interact with the Philippine state.

The framework that best describes such continuity, this article sug- gests, derives from Max Weber's work on patrimonialism. Political ad- ninistration—whether in the pre–martial law period or under martial aw or in the Aquino years—is often treated as a personal affair.[3] The tate apparatus is choked continually by an anarchy of particularistic de- nands from, and particularistic actions on behalf of, those oligarchs and ronies who are currently most favored by its top officials: one will obtain ı highly coveted loan or import license; another will enjoy a stake in a 'artelized industry unfettered by effective state regulation. As Weber vrites of the patrimonial state:

> In general the notion of an *objectively defined* official duty is unknown to the office that is based purely upon personal relations of subordination. . . . Instead of bureaucratic impartiality and of the ideal—based on the abstract validity of one objective law for all—of administrating without respect of persons, the opposite principle prevails. Practically everything depends explicitly upon the personal considerations: upon the attitude toward the concrete applicant and his concrete request and upon purely personal connections, favors, promises, and privileges.[4]

Under the martial law regime Marcos attempted to centralize a largely lecentralized patrimonial polity and thereby undercut competing cen- ers of power. The most threatening members of the old oligarchy were amed, and the rising cronies were set loose in "new field[s] of patrimo- ıialism" (Wurfel, 258). But in pursuing a program of selective expropri-

[1] For a neoclassical perspective on competitive rent seeking, see Anne O. Krueger, "The 'olitical Economy of the Rent-Seeking Society," *American Economic Review* 64 (June 1974), 91–303.

[2] Kenji Koike, "Dismantling Crony Capitalism under the Aquino Government," in Aıi- hiro Ishii et al., *National Development Policies and the Business Sector in the Philippines* (To- ·yo: Institute of Developing Economies, 1988), 255.

[3] In "the genuinely patrimonial office," Weber writes, "the political administration . . . is reated as a purely personal affair of the ruler, and political power is considered part of his ersonal property." Weber, *Economy and Society*, ed. Guenther Roth and Claus Wittich (New 'ork: Bedminster Press, 1968), 3:1028–29. Unless otherwise specified, all subsequent refer- nces are to vol. 3.

[4] Ibid., 1041; emphasis in original.

ation, creation of export monopolies, and promotion of favored associates, or "cronies," Marcos was merely expanding on earlier patterns of patrimonial plunder. Particularistic demands continued to prevail, with the difference that one ruler was appropriating a much larger proportion of the state apparatus toward the service of his own private interests. As Emmanuel de Dios puts it, "The crony phenomenon was no more than a logical extension and culmination of the premartial law process of using access to the political machinery to accumulate wealth."[5] In fact, one might say that Marcos streamlined the process: just as Weber explains that patrimonial office can "develop bureaucratic features with increasing functional division and rationalization"[6] and yet not undermine its distinctly patrimonial character, Marcos could undertake bureaucratic reform and enlist the aid of technocrats without undermining his own agenda of patrimonial accumulation.

A focus on patrimonial aspects of the state,[7] I argue, will lead to a clearer understanding of important characteristics of the Philippine political economy. It also brings a greater sense of coherence to many disparate aspects of Marco's rule, including his motivations for declaring martial law, his efforts to undercut the power bases of rival patrons, his reliance upon loyal friends and family, his promotion of crony interests and export monopolies, his extravagant use of foreign loans, and his choice of economic strategies.

This is not to obscure critical differences between the postcolonial Philippines and the economies, societies, and polities analyzed by Weber many decades ago. Most importantly, Weber's notion of historical progression does not anticipate the "neopatrimonialism" that comes in the wake of a more "rational-legal" colonial state. He likely would have con-

[5] De Dios, "A Political Economy of Philippine Policy-Making," in John W. Langford and K. Lorne Brownsey, eds., *Economic Policy-Making in the Asia-Pacific Region* (Halifax, Nova Scotia: Institute for Research on Public Policy, 1990), 114.

[6] Weber (fn. 3), 1028.

[7] My definition of the state derives from the incisive discussion of postcolonial African states found in Robert H. Jackson and Carl G. Rosberg, "Why Africa's Weak States Persist: The Empirical and the Juridical in Statehood," *World Politics* 35 (October 1982), 1–24. They go beyond Weber's classic "empirical" definition of the state ("a corporate group that has compulsory jurisdiction, exercises continuous organization, and claims a monopoly of force over a territory and its population, including 'all action taking place in the area of its jurisdiction' ") (p. 2) and argue the need to incorporate also the "juridical" definition of statehood, that which is recognized by the "international society of states." In Europe, they argue, "empirical statehood preceded juridical statehood or was concurrent with it." In Black Africa and other parts of the Third World, however, there has been a very different sort of state-building process: "External factors are more likely than internal factors to provide an adequate explanation of the formation and persistence of states" (p. 23). "A political system may possess some or all of the empirical qualifications of statehood, but without the [internationally recognized] juridical attributes of territory and independence it is not a state" (p. 13).

sidered it particularly anomalous to find patrimonial features in the former colony of the world's most advanced industrial power. As will be discussed below, patrimonial features appear to have become more pronounced *after* independence in 1946, when the economic role of the central state expanded.[8] Second, except during the Marcos years, the postwar Philippines has not had the clearly identifiable central "ruler" that Weber expected to see in patrimonial polities. Instead, a nominally strong president (see Wurfel, 76–88) of relatively short tenure must make major accommodations to "local patrimonial lords"[9] who possess economic power and assume quasi-military and quasi-judicial functions in their localities and are represented at the national level in a powerful legislature.[10]

Third, while Weber's patrimonial polities were largely self-contained, the Philippine political economy depends upon external funds for its sustenance—and Marcos's own personal strategies cannot be understood except in a world of Swiss banks and Manhattan real estate. And finally, Weber expected to see patrimonialism in economies that are not yet fully monetized and where the means of communication and transport are poorly developed. While this may partially describe some of the subnational "polities" controlled by local autonomous lords, it does not describe the national-level polity on which I am focusing. Weber would not have imagined the strengthening of patrimonial features in a "modernizing" economy. My use of the term "patrimonial features" rather than "patrimonialism" is meant to highlight the transplanting of Weber's framework into a setting that he did not anticipate.[11]

[8] This parallels Guenther Roth's discussion of a "detraditionalized" patrimonialism that "becomes the dominant form of government" in "some of the newer states." Even in industrialized countries, he writes, this "personal rulership ... is apparently enlarged by the extension of government functions"; Roth, "Personal Rulership, Patrimonialism, and Empire-Building in the New States," *World Politics* 20 (January 1968), 194–206, at 196, 199. Weber (fn. 3), 1002–3, discusses the historical "march of bureaucracy."

[9] Weber (fn. 3), 1088 and 1059.

[10] The formation and persistence of a central state that is unable to exert authority effectively over "local lords" can only be understood in light of external factors: (1) an American colonial heritage that aborted what might have been a more "natural," Latin American–style caudillo route to state formation and superimposed a weak central state over a polity of quite autonomous local centers of power; and (2) an "international society of states" that helps ensure the survival of central states that cannot, empirically, claim control over much of their territories. See Benedict Anderson, "Cacique Democracy and the Philippines: Origins and Dreams," *New Left Review* 169 (May–June 1988), 3–33, at 9–13; and Jackson and Rosberg (fn. 7).

[11] Weber (fn 3), of course, develops "pure" categories and then permits hybrid characterizations of his historical cases. In his analysis of bureaucratic systems, for example, he notes cases with "strong patrimonial elements" and "patrimonial admixtures" (p. 964). It is in this spirit that I use the terms "patrimonial features" and "patrimonial elements."

My analysis has benefited from two works that provide excellent, historically based analyses of patrimonial dynamics in other postcolonial settings: Harold Crouch, "Patrimonialism and Military Rule in Indonesia," *World Politics* 31 (July 1979), 571–87; and Thomas M. Cal-

As the following survey of current scholarship on the Marcos regin attempts to show, little attention has been paid to the patrimonial fe tures that endure—across regimes—within the Philippine state. By d veloping a clearer understanding of certain features that Weber best d scribes, we can explain more adequately how dominant econom interests have continued to plunder the Philippine state.

APPROACHES TO UNDERSTANDING THE MARCOS REGIME

Five years after the downfall of Marcos, scholars of Philippine politi have yet to achieve even minimal consensus on the proper characteriz tion of his authoritarian regime. Some see Marcos as largely an aberr tion; Belinda Aquino[12] writes that "while corruption has always bee part of Philippine political life, it reached epidemic and flagrant propo tions during the Marcos years" and displayed "a certain pathological d mension" (pp. 1, 4). Her examination of 2,300 pages of the "Marcos Pa pers" leads her to conclude that systemic explanations alone ar inadequate to understanding the "politics of plunder" under Marcos; or must also "look at the role of individual dictators themselves in the de struction of their own societies" (p. 83).

A prominent Marxist scholar, Francisco Nemenzo, sees the "Marco state" as "analogous to ... 'Bonapartism.' It achieved 'relative autonom from the ruling class with the support of the army and a pliable ma organization." Louis Bonaparte and Marcos, he argues, faced "remark ably similar" circumstances: "intense contradictions in the ruling cla and a mighty challenge from below, resulting in the paralysis of the ol state machine."[13] Similarly, business journalist Rigoberto Tiglao argue that in the early years of martial law, "the ruling elite completely capit ulated to the dictatorship," while enjoying an exceptionally profitabl period in their enterprises. He sees the origins of the Marcos regime as in part, an elite project:

> While partly representing Marcos's obsession to keep his hold on the Philippine presidency, the dictatorship would also represent the social project

laghy, *The State-Society Struggle: Zaire in Comparative Perspective* (New York: Columbia Uni versity Press, 1984). Some scholars find patrimonial features in modern First World bureau cracies as well. See the very insightful "revisionist" approach to Weber's ideal-typica bureaucratic state in Lloyd I. Rudolph and Susanne Hoeber Rudolph, "Authority and Powe in Bureaucratic and Patrimonial Administration: A Revisionist Interpretation of Weber on Bureaucracy," *World Politics* 31 (January 1979), 195–227. Robin Theobald, by contrast, seek to confine the patrimonial framework to underdeveloped economies; see Theobald, "Patri monialism," *World Politics* 34 (July 1982), 548–59.

[12] "No relation to President Corazon C. Aquino," it is noted in the foreword to her book

[13] Nemenzo, "The Left and the Traditional Opposition," in R. J. May and Francisco Ne menzo, eds., *The Philippines after Marcos* (New York: St. Martin's Press, 1985), 50.

of the dominant sections of the Philippine ruling elite to break the dead end it faced in the early 1970s. Only towards . . . the late 1970s would the elite realize the Frankenstein monster it had helped create which threatened to eat them up one by one and to drive the entire nation into an economic and political holocaust.[14]

A third school emphasizes the heightened role of external forces, particularly the United States, the International Monetary Fund, the World Bank, and foreign capital. Here, Marcos is often portrayed as acting on behalf of foreign economic and strategic interests. Robert Stauffer, for example, wrote in 1979 that "the new regime pushes the nation to accept the increasing degree of direction imposed from the agencies of the metropolitan nation and the multilateral institutions over which it has controlling influence."[15] Robin Broad follows in this dependency tradition, with a detailed examination of how the IMF and the World Bank, allied with the Marcos-supported "transnationalist" technocrats within the Philippine state, used two loans from the World Bank to outmaneuver "economic nationalists" in the government and promote export-oriented industrialization. Between 1979 and 1982, she argues, industrial and financial sector loans led to "a reshaping of the entire Philippine political economy and of the ways major institutions and social groups within it meshed" (p. 162).

Gary Hawes and David Wurfel adopt more varied approaches. Hawes argues that "martial law marks the emergence of a new, transnational coalition of class interests," and he seeks to determine exactly which class segments were dominant in the Philippine state. He sees a state that "was able to take actions that were opposed to the interests of one class segment—import-substitution manufacturers"—and that also defended "the general interests of capital." But Hawes does not try to explain Marcos entirely in terms of the transnational coalition and class segments: "Some actions taken by Marcos," he writes, "can be interpreted only in the context of domestic politics and his attempts to solidify his own regime." For example, Marcos's promotion of cronies and his creation of export monopolies was in clear opposition to World Bank, IMF, and U.S. advice. We shall see that an important part of Hawes's analysis is "the key distinction between state and regime" (pp. 52–53).

Wurfel's work is as wide-ranging in its framework as it is comprehensive and balanced as a history of recent Philippine politics. While ideas

[14] Tiglao, "The Consolidation of the Dictatorship," in Aurora Javate-de Dios et al., eds., *Dictatorship and Revolution: Roots of People's Power* (Metro Manila: Conspectus Foundation Incorporated, 1988), 27, 31 (quote from p. 27).

[15] Stauffer, "The Political Economy of Refeudalization," in David A. Rosenberg, ed., *Marcos and Martial Law in the Philippines* (Ithaca, N. Y.: Cornell University Press, 1979), 217.

of modernization, political development, and legitimacy most influence Wurfel's approach, his work contains many other ingredients as well. The analysis of the political economy under Marcos, for example, cooks up a veritable stew of theoretical terms and labels: "patrimonial authoritarianism," centralized patrimonialism," and "neopatrimonialism"; "authoritarian facism" and "corporatism"; "authoritarian-technocratic"; "neocolonialism" and "dependency"; "developmental autocrat"; "superpatron," and so on. Wurfel's purpose seems not to arrive at any single theoretical formulation but rather to provide a careful descriptive account of Filipino politics—and indeed, his book has already become the standard text on the subject. Within Wurfel's varied theoretical framework, one finds a valuable picture of the complex dynamics that characterized Marcos's Philippines: the personalistic nature of the regime, the opportunities and pressures coming from external forces, and the changes in intraelite and elite-mass relations.

In arguing the utility of the patrimonial framework, this essay examines four key aspects of recent Philippine political economy in relation to the books under review: (1) the motivations behind the declaration of martial law; (2) the role of technocrats, international multilateral institutions, and foreign indebtedness in the choice of economic strategy; (3) the rise of the cronies and the Marcos relatives; and (4) the resurgence of the "old oligarchy" in the early 1980s. As an initial step, however, it is important to delineate the critical historical characteristics of the relationship between state and oligarchy in the Philippines.

State and Oligarchy in Historical Perspective

While scholars may still debate the nature of the Philippine state under Marcos, there is little dispute that prior to 1972 Philippine state institutions were weak vis-à-vis dominant interests. As Willem Wolters explains:

> The Philippine state ... which emerged during the late-Spanish and American colonial periods, was characterized by a limited degree of centralization and a weak state apparatus. The land-owning elite in the provinces prevented the development of a strong central state. The classic state monopolies known from European history, namely those over violence and taxation, have never been fully developed in the Philippines.[16]

Unlike in Thailand and Indonesia, where "bureaucratic-aristocratic" elites (descendant of precolonial kingdoms) were strengthened by the

[16] Wolters, *Politics, Patronage and Class Conflict in Central Luzon* (Quezon City: New Day Publishers, 1984), 3.

nineteenth-century commercialization of agriculture, the same process in the Philippines "gave rise to a new class of... landowners who were quite separate from the bureaucracy."[17] Their economic base was firmly *outside* the state. From this independent base, however, they came to exercise powerful—yet particularistic—control over elements of the state apparatus through a spoils system that had become well entrenched at the national level early in this century.[18] Benedict Anderson explains that representative institutions introduced by the Americans led to the first creations of "a solid, visible 'national oligarchy,'" and through these institutions the oligarchy enjoyed suzerainty over the bureaucratic apparatus: "Civil servants frequently owed their employment to legislator patrons, and up to the end of the American period the civilian machinery of state remained weak and divided."[19] This control of the bureaucracy, it must be emphasized, was exercised from an independent societal base: while the oligarchy swamped the legislature, it showed little interest in directly assuming bureacratic posts. Despite significant growth in the bureaucracy, a bureaucratic elite never emerged.[20]

External support has been vital to oligarchic hegemony over the Philippine state and society. The consolidation of the oligarchs' position came under American colonial sponsorship, but even after 1946 both state and oligarchy remained highly dependent upon U.S. aid, investment, and counterinsurgency support. Three years after independence the Philippine state nearly collapsed, in large part because of its lack of tariff autonomy vis-à-vis the former colonial power and its inability to extract revenue from the oligarchy. The military and economic rescue operation came from Washington. As Frank Golay explains:

> By the end of 1949 the government seemed willing to let the military go unpaid and the educational system wither for want of funds, and even to succumb to the Huk rebellion, rather than face up to minimum responsibility for governmental functions.... [There was] mounting evidence that the body politic was incapable of action in the interests of all Filipinos. It

[17] Harold Crouch, *Economic Change, Social Structure and the Political System in Southeast Asia: Philippine Development Compared with the Other ASEAN Countries* (Singapore: Institute of Southeast Asian Studies, 1985), 10–18, quote at 10.

[18] See Onofre D. Corpuz, *The Bureaucracy in the Philippines* (Manila: University of the Philippines Institute of Public Administration, 1957), 249.

[19] Anderson (fn. 10), 11–12. One of the best prewar illustrations of an oligarchic raid on the state machinery is the sugar bloc's unbridled assault on the loans department of the newly formed Philippine National Bank, between 1916 and 1920. See Peter W. Stanley, *A Nation in the Making: The Philippines and the United States, 1899–1921* (Cambridge: Harvard University Press, 1974), 233–48.

[20] Daniel F. Doeppers, *Manila, 1900–1941: Social Change in a Late Colonial Metropolis*, Monograph Series No. 27 (New Haven: Yale University Southeast Asia Studies, 1984), 72. See also Paul D. Hutchcroft, "A State Besieged: Historical Patterns of State-Elite Relations in the Philippines," *Issues and Letters* (Philippine Center for Policy Studies) 1, no. 4 (1990), 1–10.

> is a depressing commentary that the reforms, when they did come, were to a considerable extent installed from the outside as a result of the Bell Mission and its recommendations.[21]

In part because the grantor of independence was a rising superpower—not a declining European power, as elsewhere in Southeast Asia—it was especially difficult for the Philippines to emerge as a truly sovereign nation. Even into the Aquino era, in fact, close relations between oligarchs and Washington policymakers endure. Throughout the postwar years oligarchs have needed external support to sustain an unjust, inefficient, and graft-ridden political and economic structure; Washington, in return, has received unrestricted access to two of its most important overseas military installations.

After independence, I will argue, certain developments led to the strengthening of patrimonial features, or a blurring of the distinction between "official" and "private" spheres.[22] Although further research is needed, we can make the following observations. First, personal contacts became even more important for entrance into the central bureaucracy, and the role of competitive examination became relatively marginal. As Wurfel reports, congressmen felt they were spending "most of their time ... running an employment agency." The bureaucracy expanded rapidly, especially at election time (pp. 80, 85).[23]

Second, in the countryside, patron-client relations were undergoing significant changes. Patrons, who historically had relied on their own local (often land-based) resources, found expanded opportunities in obtaining external and "office-based" resources. This did not diminish the power of local oligarchs vis-à-vis the central state but merely increased the role of state resources *within* long-standing patron-client relationships. In the context of national electoral system, these local oligarchs retained enormous power to milk the central state's "particularistic distributive capacity."[24]

[21] Golay, *The Philippines: Public Policy and National Economic Development* (Ithaca, N.Y.. Cornell University Press, 1961), 71–72.

[22] As Weber (fn. 3) writes, "The patrimonial office lacks above all the bureaucratic separation of the 'private' and the 'official' sphere" (p. 1028).

[23] There are two explanations commonly given for this postindependence change: (1) increased corruption in the bureaucracy during the Japanese occupation and the early postwar years; and (2) the birth of the two-party system, injecting more partisan considerations into the spoils system. See Corpuz (fn. 18), 224–25, 237–48. A third reason is perhaps most important: what Anderson calls the "American ringmaster for domestic political competition" played a less prominent role after 1946; Anderson (fn. 10), 14.

[24] James C. Scott, "Patron-Client Politics and Political Change in Southeast Asia," in Steffen W. Schmidt et al., eds., *Friends, Followers and Factions* (Berkeley and London: University of California Press, 1977), 137 and 143. Thomas C. Nowak and Kay A. Snyder explain that the "growing penetration of national bureaucratic institutions into local areas heightens the

Third, as oligarchic interests diversified (beyond agriculture, to include also commerce, manufacturing, and finance), access to the state machinery became more important than ever for the creation of wealth.[25] In the period of import and exchange controls in the 1950s, for example, rent-seeking entrepreneurs flooded the halls of the Central Bank in search of the dollar allocations that would enable them to reap windfall profits in producing for a protected domestic market. As long as such "rents" can be obtained, Manuel Montes points out, rent seekers find it more important to maintain their government connections than to concern themselves with "the internal efficiencies and investments" of their firms.[26] Indeed, one study of the textile industry in the 1950s states that entrepreneurs "considered effort at the Central Bank as important as [effort] at their plants."[27]

These postwar developments—in the central bureaucracy, local patron-client relations, and rent-seeking ventures—greatly heightened the role of the state in private accumulation and strengthened the patrimonial features of the Philippine state. Because these patterns became more pronounced in recent times, as access to the state began to be more important for securing patronage and rents, the term "neopatrimonial" best captures the historical sequence. These postwar developments occurred in the midst of important continuity in the relative strengths of state and oligarchy: whether colonial or postcolonial, the civilian state apparatus remained "weak and divided" in the face of powerful oligar-

---

dependence of the local elite upon office-based resources"; Nowak and Snyder, "Clientelist Politics in the Philippines: Integration or Instability," *American Political Science Review* 68 (September 1974), 1147–70, at 1151. See also Brian Fegan's discussion of how the closing of the land frontier made civil service positions a more important element of patron-client largesse in the postwar years—and led to a process of "bureaucratic involution" on the national level; Fegan, "The Social History of a Central Luzon Barrio," in Alfred W. McCoy and Ed. C. de Jesus, eds., *Philippine Social History: Global Trade and Local Transformations* (Honolulu: University Press of Hawaii, 1982), 119–24. An influential early work on patron-client relations in the Philippines is Carl H. Landé, *Leaders, Factions, and Parties: The Structure of Philippine Politics*, Monograph Series No. 6 (New Haven: Yale University Southeast Asia Studies, 1966).

[25] As one scholar wrote in the 1950s, "Business is born, and flourishes or fails, not so much in the market place as in the halls of the legislature or in the administrative offices of the government"; Thomas R. McHale, "An Econecological Approach to Economic Development" (Ph.D. diss., Harvard University, 1959), 217, quoted in Laurence Davis Stifel, *The Textile Industry: A Case Study of Industrial Development in the Philippines*, Data Paper Number 49 (Ithaca, N.Y.: Southeast Asia Program, Cornell University, 1963), 50.

[26] The "rent-seeking society" in the Philippines is further discussed in Montes, "The Business Sector and Development Policy," in Ishii et al. (fn. 2), 23–77. As he points out, "rents" can include "protection from competition through quotas, tariffs, and measured capacities, subsidized credit, access to foreign reparations, loans, and grants. . . . [I]n a rent-seeking society, the operations of the state determine the assignment of and the continued enjoyment of economic advantages" (p. 65).

[27] Stifel (fn. 25), 104.

chic interests. Moreover, even as raids on state resources increased in importance, the oligarchy retained its firm economic base *outside* the state.

Any serious analysis of the Philippine state, then, must paradoxically begin *outside* the state: first, because the patrimonial nature of the system greatly blurs the distinction between "official" and "private" that is found in Weber's genuinely bureaucratic state; and second, because a generally incoherent bureaucracy (with a poorly developed bureaucratic elite) has rarely been able to counter the power of the oligarchy over the state apparatus. Put another way, in relations between the state and the oligarchy, the state is far more likely to be *acted upon* than to be an *independent actor*. For this reason, we must understand the nature of the dominant interests that pursue their particularistic agendas within the Philippine state.[28]

The Declaration of Martial Law

A focus on dominant interests is particularly important for a discussion of the declaration of martial law. Hawes's explantion rests largely upon an analysis of "elite segments" and the transition to a new model of development. "Elite cohesion broke down during the 1960s," he argues, as the result of the increasing power of import-substitution industrialists, "the influx of foreign investors," and the heightened participation of IMF, the World Bank, and the multinationals. "One segment, broadly nationalist . . . , favored protection for domestically owned industry and curbs on foreign investment," while the other segment favored "an open economy" (pp. 36–37). Marcos, he argues, was in the second camp, and with the declaration of martial law in 1972, he "resolved the breakdown in

[28] Theda Skocpol writes that "states conceived as organizations claiming control over territories and people may formulate and pursue goals that are not simply reflective of the demands or interests of social groups, classes, or society. This is what is usually meant by 'state autonomy.' Unless such independent goal formulation occurs, there is little need to talk about states as important actors"; Skocpol, "Bringing the State Back In: Strategies of Analysis in Current Research," in Peter B. Evans et al., eds., *Bringing the State Back In* (Cambridge: Cambridge University Press, 1985), 9.

While it may be rare to speak of the Philippine state as an independent actor vis-à-vis dominant interests, the state nevertheless remains central to any comprehensive analysis of the country's political economy. As noted at the outset, access to the state apparatus is the primary means of private accumulation. It is indeed paradoxical that a "weak" state should be a central subject of analysis (and, as noted above, that analysis of the state must begin outside the state). The state's important role seems to derive primarily from responsibilities it has necessarily assumed in handling the country's *external* economic relations: it disburses aid and loans received from abroad and sets policies on foreign exchange, trade, and investment. In such a paradoxical situation, we must *begin* our analysis with an examination of the nature of dominant interests and proceed from there to develop a clearer picture of the state with which they interact.

cohesion" (p. 20) and "cleared the way for ... export-led industri-tion" (p. 45). Wurfel seems at points to concur with this analysis (pp. o, 192, 238).

number of problems emerge from this interpretation of events. , if Hawes is correct that the state took actions against the import-:itution industrialization (ISI) manufacturers and was able to "punish ıents of the bourgeoisie opposed to ... export-oriented industriali-n [or EOI]" (p. 14), we should be able to see this in the choice of :ts after the declaration of marital law. In fact, although outspoken ınalists such as Senator Jose Diokno were arrested, there is no evi-e of a concerted attack against businessmen who produced for the :stic market. In true patrimonial fashion, Marcos's targets after the ration of martial law were those rival clans who threatened his ehold, *not* "elite segments" standing in the way of a new approach onomic development.[29] While Marcos claimed to be attacking the rchs, his attacks were actually limited to relatively few of their num-including "his two most potent competitors, Benigno Aquino and r baron' Eugenio Lopez" (Wurfel, 21). As historian Alfred McCoy .ins, Marcos

> *dified but did not alter the familial basis of business and politics.* ... With ısiderable dexterity, Marcos employed armed force, masked with law, ɔursue the main aim of his rule—changing the composition of regional l national elites.[30]

ıe second major problem with Hawes's analysis is that it posits elite ages *based on clear differences of economic interest*. Such cleavages are, :t, difficult to find. Stauffer tried to locate them but could not. After ng the journals of the major business associations of this period, he udes that while individual businessmen supported nationalist s, "none of the business associations supported the nationalists. ... :se groups largely accepted the need to rely heavily on foreign loans nvestments."[31] Third, ISI was not undermined by martial law, and

ıen Hawes seems to acknowledge this, despite the apparent inconsistency with his about "the state" acting against "elite segments": "Occasionally individual families ingled out and made an example so that others might see what fate awaited those d to be enemies of the president" (p. 127).

cCoy, "The Lopez Family: From Provincial Elite to National Oligarchy," in Michael ıne, ed., *Philippine Political Families* (Ann Arbor, Mich.: Center for South and South-ıia, University of Michigan, forthcoming); emphasis added.

auffer (fn. 15), 193. Behind Hawes's "elite segments" seems to be the notion that pro-ı for the domestic market (or import-substitution industrialization) is generally na-ıt, while production for external markets is not (see p. 45). The difficulty here is that ı capital has played a central role in Philippine ISI from the beginning (which Hawes y acknowledges, p. 33). See Sylvia Maxfield and James H. Nolt, "Protectionism and ernationalization of Capital: U.S. Sponsorship of Import Substitution Industrializa-

EOI did not emerge triumphant. As shall be discussed below, the strategy was one of "debt-driven growth";[32] as long as the system was being lubricated by external funds, there was no need to make any hard decisions regarding contending economic paths.

Resolving the difficulties in Hawes's analysis requires a clearer understanding of the basis of intraelite competition in the pre–martial law years. The basic building blocks of the political economy of the Philippine oligarchy are not "elite segments" but extended families. Hawes recognizes the particularistic nature of the system, when he speaks of the "tendency of the state to respond in preferential, partisan fashion to demands from segments of the bourgeoisie" (p. 32). But the state is much more likely to be responding to families than to any other coherent segments. Philippine business associations are notoriously weak and poorly institutionalized, and its members know that the way to make money is to gain privileged access to the government and then "to exclude information from each other."[33]

Both Hawes and Wurfel acknowledge the importance of family groupings, though they nonetheless persist in arguing the presence of other definable elite segments. Hawes admits that "division of the bourgeoisie may be more important analytically than in real life, because many of the major families of the Philippines have economic empires

---

tion in the Philippines, Turkey and Argentina," *International Studies Quarterly* 34 (March 1990), 49–81.

Moreover, Montes points to the commonality of interest between Filipino firms and foreign firms that are protected by the same "tariff walls and protectionist measures"; Manuel F. Montes, "Financing Development: The 'Democratic' Approach versus the 'Corporatist' Approach in the Philippines," in Miguel Urrutia et al., eds., *The Political Economy of Fiscal Policy* (Tokyo: United Nations University, 1989), 98.

Interestingly, if one is looking for coherent segments of capital based on different economic interests, it can best be found in analyzing *foreign* capital. There, Montes points out, one can see a clear differentiation between foreign capital inside the tariff walls and foreign capital located outside the country (p. 98).

[32] Montes (fn. 31), 90.

[33] Interview, Wilhelm G. Ortaliz, former director, Bureau of Industrial Coordination, Ministry of Industry, April 26, 1989. Ortaliz explained that his job during the early 1980s was to "strengthen industrial associations within the private sector.... [B]ut it was hard for the government to know who to work with." He described the largest business association, the Philippine Chamber of Commerce and Industries (formed in the late 1970s), as a mere "post office" of diverse concerns, very "personality-oriented" and unable to formulate common positions on major issues.

Economist Montes similarly explains that government–private sector "consultations are dominated by the need to protect individual interests.... The private sector does not have a well-defined interest as a private sector (or as larger groups within the private sector). This protecting of individual interests also involves the need to have advance information of the government's intentions.... Access to information tends to be individual, based on kinship and other ties"; see Montes (fn. 31), 108.

Indeed, one can say that quasi-patrimonial features (mirroring those found at the level of the state) are found within business associations, as personal interests usually win out over associational interests.

that transcend these analytical categories" (p. 135; see also Wurfel, 57). In real life Philippine firms tend to be closely held within diversified family groups, many of which include a bank whose loan portfolio can be milked by family enterprises. Loyalty is primarily to the family grouping, not to any of its agricultural, industrial, commercial, real estate, or financial elements.[34]

The primary tensions late in the pre–martial law period, then, have little to do with the "elite segments" described by Hawes. De Dios provides the best analysis of the situation:

> By 1970 . . . three sets of circumstances combined to make matters come to a head. The first was a worsening economic crisis in the form of severe balance of payments difficulties provoked by heavy electoral spending during the presidential elections. . . . The resulting inflation ate into real incomes . . . and led to worker-student unrest. . . . [Second,] the uncertain status of U.S. assets in the country, owing to the impending termination in 1974 of parity rights given to U.S. citizens. . . . [And third was Marcos's desire for a third term as president.] The split in the Filipino elite had reached crisis proportions, owing mainly to this bid for political and financial hegemony by the Marcoses and Romualdezes (who were relative upstarts) as against the more established political clans such as the Lopezes, Osmenas, Aquino-Cojuangcos, and Jacintos.[35]

There *was* intensifying intraelite competition, but it manifested itself largely along clan lines. Wurfel writes that each biennial campaign in the 1960s led to "a few hundred political deaths" (p. 103). Vote buying was rampant, and "private armies" were a central component of many campaign organizations. The 1969 elections were a turning point in the developing crisis, as all candidates (led by Marcos, who raided the public treasury) spent the equivalent of nearly one-fourth of the national budget (Wurfel, 100). As Marcos's defeated opponent grumbled, "[We were] out-gooned, out-gunned and out-gold."[36]

The economy simply could not support the political system. Consis-

[34] Nowak and Snyder (fn. 24) explain that "diversification of familial economic power decreases susceptibility to fluctuations in world prices, potential loss of privileges in the U.S. market, and local policy changes such a devaluation, which hurt some sectors more than others" (pp. 1148–49). They also note the important, but often neglected, role of urban real estate in diversification strategies.

There is one other reason for diversification that should be added to their list: in an economy in which wealth depends to such a large degree on access to the state machinery, diversification helps to guard against the uncertainties of change in political leadership. A family cannot depend exclusively on investments assisted by *current* friends in the Palace, for example, because in the *next* administration those investments may be jeopardized by a lack of necessary connections in key government offices.

[35] De Dios (fn. 5), 112.

[36] Jose Veloso Abueva, "The Philippines: Tradition and Change," *Asian Survey* 10, no. 1 (1970), 56–64, at 62.

tent with James Scott's model of "patron-client democracy," the state had a weak revenue base but was nonetheless faced with "intense distributive pressures ... especially in election years."[37] Despite increasingly scarce national patronage resources, there was a proliferation in the number of local candidates.[38] Moreover, the economy was stagnating: the "easy phase" of ISI had long been exhausted, and export growth had been unimpressive for the overall period since decontrol (of imports and exchange) in 1962.[39] The result was a major balance of payments crisis in the wake of the 1969 elections, followed by a devaluation that fueled inflation. This, in turn, heightened popular demands for change.[40] In a period of constricted opportunity, the oligarchs became especially vicious in clawing for the booty of state. "American-era inhibitions slackened," Anderson notes, and "it was only a matter of time before someone would break the rules and try to set himself up as Supreme Cacique for Life."[41]

External Resources and the Agenda of Economic Reform

In the short term martial law resolved the overload on the state's "distributive capacity": the abolition of Congress eliminated expensive intra-elite electoral competition, and both heightened repression and the curtailment of elections restricted demands from the masses.[42] As will be discussed below, Marcos attempted to centralize a hitherto decentralized patrimonial polity and thereby achieve tighter control over access to the state apparatus. At the same time he worked on the supply side problems of the system's overload and obtained new resources internationally. This was "the most important task" of the regime's foreign policy, remarks Wurfel (p. 190), and for a full decade Marcos adroitly managed

[37] Scott (fn. 24), 143. The revenue problems are systemic:

> A regime that is dependent on its particularistic distributive capacity ... [will] have a most difficult time raising revenue from internal taxation. A rise in direct taxation would threaten their base of support, and in fact, they are notorious for the undercollection of revenues due them, since favors to their clients often take the form of either leaving them off local tax rolls or ignoring debts they owe the government. (p. 143)

[38] Nowak and Snyder (fn. 24), 1151–54.

[39] See Romeo M. Bautista, John H. Power, and Associates, *Industrial Promotion Policies in the Philippines* (Metro Manila: Philippine Institute for Development Studies, 1979), 5–9; and Montes (fn. 31), 88–89.

[40] As Scott (fn. 24) explains:

> Democratic regimes which must cater to the strong distributive pressures generated by their clientele are thus particularly vulnerable to the vagaries of world prices for primary products on which their budgets depend.... [A] stagnating economy or declining world prices threatened the entire structure they had pieced together. (p. 143)

[41] Anderson (fn. 10), 18.

[42] See Nowak and Snyder (fn. 24), 1170.

enormous sums from the IMF and World Bank, bilateral do- commercial banks. The system was lubricated once again.
began martial law with the good fortune of high international ity prices, which "generated windfall profits for the Philippine c elite, dispelling whatever doubts it still had about the Marcos hip."[43] Later in the decade the regime took full advantage of the lity of cheap petrodollars at negative real rates of interest.
ood fortune alone cannot explain the regime's success. Two gen- s of Filipino oligarchs had tapped American patronage to boost ersonal positions domestically,[44] but among postwar leaders Mar- played particularly keen insights into the nature of the neocolonial As a man whose primary loyalty was to personal interests rather any state or national interests—however defined—Marcos saw merican strategic needs presented ample opportunity for private long history of American manipulation of Philippine affairs not- anding,[45] Marcos knew that neocolonial manipulation can be a ay street.[46] "More clearly than anyone else," Anderson explains, he stood "that for Washington the Philippines were like Cyprus for n."[47] Especially at a time when the military bases were offering mportant support to U.S. forces in Vietnam, Marcos could ap- Washington aid givers from a position of strength. Indeed, the ewarded martial law with very large increases in grants and loans el, 191).
the same time close relations with the U.S. assisted Marcos in his to cultivate close relations with the IMF and the World Bank.[48] ought a corps of technocrats into his government and promulgated s of reform agendas that ensured the steady flow of multilateral ommercial bank loans. In hosting the annual IMF/World Bank g in Manila in 1976, the regime apparently dazzled its guests. nto the early 1980s, explains an American banker, "Filipinos had al hold on these institutions.... Cesar Virata [minister of finance

lao (fn. 14), 38.
enlightening scholarship on this process, see Ruby R. Paredes, ed., *Philippine Colo- ocracy*, Monograph No. 32 (New Haven: Yale University Southeast Asia Studies,

a particularly well documented account of this history, see Stephen Rosskamm Sha- *United States and the Philippines: A Study of Neocolonialism* (Philadelphia: Institute tudy of Human Issues, 1981).
mond Bonner gives a good account of how the Marcoses maneuvered in Washington ee Bonner, *Waltzing with a Dictator: The Marcoses and the Making of American Policy* ork: Times Books, 1987).
lerson (fn. 10), 21.
han Haggard, "The Political Economy of the Philippine Debt Crisis," in Joan Nel- *Economic Crisis and Policy Choice: The Politics of Adjustment in the Third World* on: Princeton University Press, 1990), 219.

and, after 1981, also prime minister] and other Marcos technocrats were intensely admired figures."[49] In assessing this cozy relationship, the basic question that observers debate is whether Marcos ever genuinely adopted the reform agendas of the World Bank, the IMF, and his technocrates, *or* whether he instead merely used their agendas as a means of aggrandizing his own regime. There is no question that he gave the Pentagon full run of the bases—but did he give the multilateral institutions and the technocrats full run of the economy?

Broad, in her analysis of two World Bank loans made between 1979 and 1982 (a structural adjustment loan for the industrial sector and an apex loan for the financial sector), stresses a unity of interests between Marcos and "the corps of Western-educated technocrats who underpinned [him]" (p. 16). She also asserts the strength of the technocrats and the World Bank within the Marcos regime, when, for example, she writes: "As early as mid-1981, Marcos announced, 'I'm going to sit back and let the technocrats run things' as a follow-through on World Bank suggestions that its favored technocrats be accorded hegemonic positions—political as well as economic" (pp. 176–77). But it is a mistake to take at face value—as Broad apparently does—Marcos's public statements of deference to the technocrats.[50] Even if we know that Marcos *was* trying to please international creditors, it is incorrect to assume that he was their lapdog.

Broad carefully chronicles the efforts of the World Bank and the IMF to orient Philippine policy-making toward the export of manufactured goods, and she makes an important contribution in drawing attention to the critical role that these institutions play in the formulation of domestic policy.[51] Unfortunately, however, she relies excessively on one key demarcation as she tries to explain the dynamics of economic policy-

[49] William H. Overholt, "Pressures and Policies: Prospects for Cory Aquino's Philippines," in Carl H. Lande, ed., *Rebuilding a Nation: Philippine Challenges and American Policy* (Washington, D.C.: Washington Institute Press, 1987), 98.

[50] Two years later, in fact, Marcos declared that the "KBL [New Society Movement, Marcos's political party] Central Committee, not the technocrats, make government policy" (Wurfel, 237 n. 8). The central committee included major cronies.

[51] Broad explains at the outset that her primary theoretical reliance is upon the "new *dependentistas*," especially Peter Evans's *Dependent Development*, but Evans's influence is not clearly evident in her analysis. The major weakness of earlier dependency analyses, she writes, "was the derivative position in which it placed the LDC state" (p. 16). But when all is said and done, Evans's "incontrovertible" assertion on "the centrality of the state to accumulation on the periphery" (p. 43) seems to have lost out to a focus more like that of Cheryl Payer, on how external institutions manage to overpower Third World states. See Evans, *Dependent Development: The Alliance of Multinational, State, and Local Capital* (Princeton: Princeton University Press, 1979); Payer, *The Debt Trap: The International Monetary Fund and the Third World* (New York: Monthly Review Press, 1974); and idem, *The World Bank: A Critical Analysis* (New York: Monthly Review Press, 1982).

making during the Marcos years. This is the struggle of the "transnationalist" faction of the state and of private interests *versus* the "nationalist" faction of the state and of private interests. Unlike the framework of Hawes (and especially that of Wurfel and Aquino), her framework downplays any independent role for Marcos's own regime interests. As Broad states her approach, "Policy formulation is viewed as emanating from the interaction of interests of local transnationalist classes and international institutions, challenged, with varying degrees of success, by nationalist factions" (p. 19). This greatly understates the influence of cronyism on the economic decision making of the martial law regime. While she does not ignore the phenomenon, neither does she successfully incorporate it into her framework.

The Central Bank, for example, is described as "the locus of nationalist power in the Philippine government" (p. 128), because of its resistance to EOI. During the 1970s, she argues, it had successfully resisted IMF pressure for import liberalization, primarily for two reasons: first, because of "allegiance to national entrepreneurs protected by import restrictions" and second, because restrictions on imports increased its power. The first reason fits squarely within her theoretical framework, but the second raises questions as to how important "nationalist" considerations actually were in explaining the intransigence of the Central Bank. The head of the import control office, in fact, was Marcos's relative; "around [her]," explains one economist, "grew a host of rent-seeking anecodotes."[52]

Broad's greatest problem is characterizing the major cronies. She tries to put them on one side of the divide:

> Among the most vocal transnationalists—and the biggest recipients of foreign bank loans—were Marcos's so-called cronies, especially Herminio Disini, Eduardo Cojuangco, Jr., and Ricardo Silverio. In addition to being heavily favored in government contracts, these men created empires in large part through transnational ties. Disini netted millions of dollars from Westinghouse in his role as agent for the Bataan nuclear-power-plant deal. Silverio made it big by winning the Philippine franchise to assemble Toyota cars. Cojuangco, who created a virtual coconut monopoly, depended on diverse transnational links to market his coconut products abroad. (pp. 106–7)

On closer examination, their loyalty to transnationalism is difficult to show. Aquino explains that Disini (who married a cousin of First Lady Imelda Marcos and was one of the president's golfing partners) benefited

[52] Raul V. Fabella, "Trade and Industry Reforms in the Philippines: Process and Performance," in Manuel F. Montes and Hideyoshi Sakai, *Philippine Macroeconomic Perspective: Developments and Policies* (Tokyo: Institute of Developing Economies, 1989), 193.

from a 1975 presidential decree imposing a 100 percent import tax o tobacco filters for all companies *other* than Disini's Philippine Tobacc Filters Corporation. His own firm was subject to only a 10 percent ta "This differential effectively shut down Disini's competitors, [including ... a company owned by American and British interests," and enable Disini to corner 90 percent of the local market for tobacco filters (p. 45 Like Disini, Silverio was basically an ISI entrepreneur, producing car for the domestic market—behind tariff walls that the IMF and Worl Bank were trying (without much success) to tear down. Cojuangco' United Coconut Oil Mills bought out Cargill's two subsidiaries in th Philippines in 1979 (Hawes, 76), and his coconut monopoly faced con certed opposition from the multilateral institutions in the final years o the Marcos regime. Broad acknowledges this pressure but asserts tha Cojuangco "would agree with the Bank ... on most other points" (p 107).

This is not to argue that the cronies are necessarily "nationalist. Rather, the intent is to highlight the serious problems with Broad's cen tral demarcation of Philippine state and society. The cronies made thei millions primarily because their intimate ties to the Palace opened u an array of opportunites, some involving close links to foreign capita and some involving the takeover and displacement of foreign firms Their loyalty was *not* to any particular economic strategy, but to Marco and the "rent-seeking orgy"[53] over which he presided. As Weber writes

> *The patrimonial state offers the whole realm of the ruler's discretion as a hunting ground for accumulating wealth.* Wherever traditional and stereotyped prescription does not impose strict limitations, *patrimonialism gives free rein to the enrichment of the ruler himself, the court officials, favorites*, governors, mandarins, tax collectors, influence peddlers, and the great merchants and financiers who function as tax-farmers, purveyors and creditors. *The ruler's favor and disfavor, grants and confiscations, continuously create new wealth and destroy it again.*[54]

Broad actually admits that "the categories nationalist and transnationalist are not strictly demarcated; numerous entrepreneurs operated in both realms" (p. 107).[55] Indeed, Disini was likely just as happy with the

[53] Ibid., 205.

[54] Weber (fn. 3), 1099; emphasis added.

[55] Problems with Broad's nationalist/transnationalist categorization arise not only with the cronies, of course, but also with noncrony businesspeople. Except for this one acknowledgment that the categories are "not strictly demarcated," she generally seems to suggest that mere involvement with foreign capital taints local capital and "tends to wed them ideologically to policies furthering free international flow of goods and capital" (p. 7). Interestingly, one of her key examples of a besieged economic nationalist is Hilarion Henares, who she says "professed to have kept his hands clean of involvement with foreign corporations and foreign

windfall profits he enjoyed as a "national entrepreneur" displacing foreign firms in the tobacco filters market as he was with the "transnationalist" kickbacks he received from Westinghouse (Aquino, 59–62). For him, what mattered was that he was a "favorite." The TNCs, too, seemed to know the best way to enter the "hunting ground"; as Broad reports, "The cronies were the most logical choice for TNCs' partners in joint ventures" (p. 171).

Ultimately, Broad has to fall back on a crony/noncrony distinction in her discussion of the bailout of crony firms after the financial crisis of 1981—an effort that fueled enormous resentment among noncrony businesspeople. "The financial squeeze," she writes, pushed "transnationalist cronies closer to Marcos while pulling non-crony transnationalists into the Marcos opposition" (p. 229). One "transnationalist crony" discovered that having ties to Toyota meant little once the "ruler's ... disfavor" descended upon his house. For seemingly personal reasons, the Palace abandoned Silverio after the collapse of his industrial and financial empire in the early 1980s.[56]

Broad does recognize certain issues that divided Marcos from "transnationalist" forces, specifically tariff reform (p. 90) and import liberalization (p. 62). But one study, examining over a decade of IMF credit to the Philippines, notes a *consistent* pattern in which "the 'dictates' of the IMF were often circumvented or simply ignored by the Philippine government." Two separate agendas are evident: "While the chief objective of the IMF was to effect short-term and structural adjustments ..., it seems the government's aim was to increase foreign inflows."[57] The IMF-sponsored tariff reduction program, for example, was undermined in part because specific corporations were exempted through presidential decree; similarly, efforts to increase the country's tax effort were hindered in part by tax incentives granted to Marcos associates.[58] Economist Raul Fabella explains further:

markets" (p. 115). In fact, Henares has long had close linkages with U.S. capital, though that seems to have had little impact on the nationalist views he regularly expresses in his popular newspaper column. Henares's business interests are discussed in William J. Pomeroy, *An American Made Tragedy: Neo-Colonialism and Dictatorship in the Philippines* (New York: International Publishers, 1974), 58.

[56] Some speculate that Silverio's fall from grace was related to displeasure over the course of a romance between Silverio's son and Imee Marcos, the daughter of the First Couple. As Weber (fn. 3) writes: "Because of its very nature patrimonialism was the specific locus for the rise of favoritism—of men close to the ruler who had tremendous power, but always were in danger of sudden, dramatic downfall for purely personal reasons" (p. 1088).

[57] Mark Thompson and Gregory Slayton, "An Essay on Credit Arrangements between the IMF and the Republic of the Philippines: 1970–1983," *Philippine Review of Economics and Business* 22, nos. 1 and 2 (1985), 59–81, at 60 and 78.

[58] Montes (fn. 31) 105, 110–11, 134.

Technocrats were given the prerogative to formulate and rhetc public agenda in the form of economic and development plan formed the basis for foreign loans. The political leadership then the unconstrained introduction of exceptions that made complete ı of the spirit and letter of the plans.[59]

The World Bank and the IMF were more comfortable with nocrats but ended up providing enormous support to the crony On the whole, Marcos usually managed to "take the money . As former Central Bank Governor Gregorio Licaros told Brc more money they [the Bank and Fund] bring in here, the bette *Certain elements of the externally induced reform agendas were in the early 1980s, when the scarcity of funds on internation. markets increased the leverage of multilateral institutions.*[61] *B means was there "a reshaping of the entire Philippine politic omy," as Broad claims.* First, even in the depths of the financ after 1983, Marcos successfully resisted the dissolution of the exp monopolies (Wurfel, 292). Second, as long as external funds were available, there was never any need to make a comprehensive | export-oriented industrialization; it was much more expedient si adopt the strategy of "debt-driven growth." In any case, the ' oriented strategy . . . did not represent the interests of any sociall icant class,"[62] and the technocrats lacked clout. Over time it bec creasingly evident that "the agenda was ultimately set by busir political interests closer to the Palace."[63]

The Rise of the Cronies and the Relatives

These Palace interests, of course, were primarily Marcos, his cror his relatives. The cronies were "a mixed bag, including not only ı of Ferdinand and Imelda, but favoured oligarchs and quite a fe

[59] Fabella (fn. 52), 197.

[60] As de Dios writes: "There were many instances when the more 'irrational,' 'ir at times blatantly corrupt, aspects of the dictatorship were countenanced or accon by these institutions, particularly its net lending operations and crony bail-outs"; "The Erosion of the Dictatorship," in Javate-de Dios (fn. 14), 122.

[61] See Mario B. Lamberte, "Financial Liberalization: What Have We Learnt?" *Philippine Development* 12, no. 2 (1985), 274–89; and Florian Alburo and Geoffrey S "Trade Liberalization Experience in the Philippines, 1960–84," Working Paper] (Manila: Philippine Institute of Development Studies, 1986). While both studies ci tant initiatives in the early 1980s, they also show that liberalization efforts ultim little success. Lamberte diagnoses the "failure of recent liberalization efforts," whi and Shepherd explain that the "1983 economic crisis . . . effectively aborted these fu tal changes in tariff structure and commercial policy" (p. 37). See also Montes (fn. 11.

[62] De Dios (fn. 60), 119–20.

[63] De Dios (fn. 5), 115.

men.' "[64] Disini is the classic new man, a virtual unknown prior to acquiring a near monopoly of the tobacco filter market. By contrast, cronies Roberto S. Benedicto and particularly Cojuangco emerged from solidly oligarchic backgrounds.

"Crony" is used to describe those whose positions are particularly favored by the current regime, regardless of their origins. An "oligarch" may not be a current crony but in either case has already established his or her fortune in earlier dispensations.[65] The Lopezes are an example of noncrony oligarchs; they depended heavily upon the state in the creation of their diversified empire, but since they were not dependent upon any particular regime, it is probably not proper, at any point, to describe them as "crony."[66] Upon the declaration of martial law, Marcos unleashed a rhetorical attack on the "oligarchs," but, as noted above, his targets were selective. In the course of his regime, both "old oligarchs" and "new men" gained "crony" status, and they were sometimes referred to collectively as the "new oligarchy."

A degree of "social mobility," or dynamism in elite composition, is an important characteristic of cronyism and underlines the patrimonial nature of state-oligarchy relations in the Philippines. As Weber explains:

> Patriarchal patrimonialism is much more tolerant than feudalism toward social mobility and the acquisition of wealth. The patrimonial ruler does not like independent economic and social powers. ... [B]ut he also does not support status barriers.[67]

Because of the steady stream of those with newfound wealth, therefore, it is difficult to reduce the Philippine oligarchy to a certain number of old families. Each presidency can create a new set of cronies, and a presidency as qualitatively different as that of Marcos can be expected to have a particularly strong influence on elite composition.[68]

[64] Anderson (fn. 10), 22.

[65] Wurfel thus confuses matters when he writes of "oligarch and crony" as an "elite cleavage on economic issues" (p. 238). A clearer way to express this division is "favored" (crony) versus "nonfavored" or "less-favored" (businessperson).

[66] McCoy explains that the Lopezes achieved national prominence after independence, building on prewar provincial wealth that was liquidated during the war. In both the pre- and postwar phases of their accumulation, "the viability of all Lopez investments ... has depended upon state regulation and licensing.... Don Eugenio [Lopez] used capital to secure political protection, investing in political campaigns and taking his profits in favors"; see McCoy (fn. 30), 21–23. In effect, they raided the state from a strong and independent societal base.

[67] Weber (fn. 3), 1102.

[68] Marcos had unprecedented tenure in office and managed to grab control of military, judicial, and legislative functions that had previously been dispersed among many rival centers of power. The measure of protection previously available to those oligarchs "on the outs" therefore disappeared in 1972. See Anderson (fn. 10), 12, 22, on how the declaration of martial law upset the "rule of law" that had long guided the system of oligarchic hegemony.

The basis of crony selection was not entrepreneurial skill, of course, but loyalty and long-standing personal connections.[69] Benedicto was a fraternity brother, Rodolfo Cuenca and Disini were golfing buddies (Aquino, 39), and Benjamin Romualdez was a brother-in-law. Marcos and Cojuangco shared an enmity with the rival side of the Cojuangco clan (that of Corazon Cojuangco Aquino and her husband Benigno) that seems to have forged a close and enduring personal bond. To the extent that the cronies were "new men," or to the extent that they relied exclusively upon Marcos for state benefits necessary to open up a particular "hunting ground," the crucial patrimonial element of loyalty and "personal dependence" was largely assured.[70] Perhaps as part of Marcos's efforts to curb any tendency toward independence, few cronies were able to combine their access to officially granted privileges with the formal assumption of official positions.[71]

By far the greatest strides toward independence were achieved by the First Lady, Imelda Romualdez Marcos, who was concurrently minister of human settlements, governor of Metro Manila, and roving diplomat (often on quite sensitive missions). Palace insiders, Wurfel writes, described the First Couple's relationship "as a standoff between two warring camps, in which the battlegrounds were appointments, government contracts, investment opportunities, media treatment, and priority in the allocation of funds" (p. 241). Whatever the precise origins of her power,[72] there is no question that at some point after 1972 she was able to establish a relatively autonomous power base within the regime. In Weber's words, the "boundaries" of her official positions were "frequently indeterminate," and the positions themselves were treated "as a *personal* right." Among government agencies, much of the jurisdictional limitation of the powers of office came from "the competing economic interests

[69] The First Lady had a special explanation for the success of cronies and relatives: "Some are smarter than others" (Wurfel, 237).

[70] On the importance of "personal dependence" to a patrimonial ruler, see Weber (fn. 3), 1026–27.

[71] For example, Cojuangco, Disini, Silverio, Rodolfo Cuenca (the construction magnate), and Antonio Floirendo (the "banana king") did not hold important official posts. Benedicto headed the quasi-public Philippine Sugar Commission, the central agency of the sugar monopoly. Major exceptions to the rule include Geronimo Velasco (energy minister) and Juan Ponce Enrile (defense minister). By the early 1980s Benedicto, Cojuangco, and Floirendo were made regional party chairmen of Marcos's *Kilusang Bagong Lipunan*, or New Society Movement (Wurfel, 237). See also de Dios (fn. 5), 114.

[72] According to a former Marcos adviser, Imelda would threaten Marcos with exposure (of intimate details of life in the Palace) or a "big public divorce, and Marcos never wanted to call her bluff." Interview, Adrian Cristobal, former Special Assistant for Special Studies (to President Marcos), June 19, 1989. Wurfel describes it as a relationship of "mutual blackmail" (p. 241). An embroidered analysis of the relationship is found in Sterling Seagrave, *The Marcos Dynasty* (New York: Harper and Row, 1988), chap. 10.

of the various patrimonial officials."[73] In this competition, the First Lady often emerged triumphant. As Montes explains:

> It became normal practice ... for the president to require that ministries attain a 10 percent savings on their current budget. The funds that were thereby saved were used for special projects, notably those of the First Lady. The affected ministries learn of the reasons for such restrictions only informally.[74]

Some observers have tried to determine whether Marcos promoted cronyism at the outset or whether the regime only degenerated into cronyism at a later stage. Wurfel, for example, characterizes the early martial law years as "authoritarian-technocratic," which by the late 1970s had been gradually replaced by a "neopatrimonial" mode (pp. 325, 331). At another point, however, he writes that there was "a slight [and temporary] shift ... toward the technocrats in 1981," in part because Virata was named prime minister (p. 254).

As Wurfel shows, it is very difficult to separate out technocratic goals and patrimonial goals. Commenting on Wurfel's thesis, Marcos's chief political theorist Adrian Cristobal remarked that "the intent was basically the same all along. You build on your success." Consistent with Belinda Aquino's many accounts of regime plunder, he noted that the cronies "had been with Marcos from the very start. They were just waiting for the opportunity." Cristobal, however, explained that Wurfel's "'authoritarian-technocratic' mode" was also present from the beginning: Marcos was "always receptive to good ideas," such as administrative reform, the development of bureaucratic capacity, and more rational development planning. In short, Marcos "believed he could have a vision for society ... and still loot it. I do not find that irrational."[75]

Technocrat-crony disputes notwithstanding, Wurfel is likely overstating matters when he describes the regime as "an *unstable mix* of patrimonialism and rule by technocrats" (Wurfel, 325; emphasis added). In fact, Marcos was generally very adept at managing the admixture. He could promote technocrat Virata and yet ensure the primacy of his own

[73] Weber (fn. 3), 1029; emphasis in original. He writes that "this quasi-jurisdictional limitation of the powers of office results primarily from the competing economic interests of the various patrimonial officials."

[74] Montes (fn. 31), 108.

[75] Cristobal (fn. 72). Similarly, an (anti-Marcos) professional technician for the state-owned Philippine National Oil Company, decries the return of unbridled patronage within his company since 1986 and looks back with an ironic sense of fondness for what one Marcos crony, Geronimo Velasco, did to professionalize their operations. Unlike the leadership under Aquino, where professional qualifications matter little and "the envelope with a recommendation from a senator is key," he exclaimed, "at least Velasco built up something worth plundering!"; anonymous interview, April 2, 1989.

patrimonial objectives.[76] As a former Palace insider explains: Marcos would utilize the skills of competent officials but at the same time try to limit their "political clout" and keep them "segmented ... in their fields."[77] In the same way that Weber allows for a bureaucratic rationalization of patrimonial leadership" that does not undermine the essential nature of that leadership,[78] one can say that Marcos's early efforts at bureaucratic reform (Wurfel, 135–38) only streamlined his plunder of the state. An important consideration in the relative portions of the "mix," it seems, is a regime's security of tenure; if there is a feeling that the regime will endure into the long term, there is no necessity to maximize gains in the short term. After the declaration of martial law Marcos likely felt secure enough in his position to know that he, personally, would be able to reap the benefits of a better-run state apparatus.

If there was simultaneous bureaucratic reform *and* patrimonial plunder, how is one to understand the concepts of state and regime? Hawes makes a valuable contribution in asserting "the necessity of clearly differentiating between the state and the regime" in the Philippine case (pp. 152–53). Alongside his problematic definition of the class basis of the Philippine state, he demonstrates very clearly how "the Marcos regime used the powers of the state to further its own economic and political interests" (p. 134). Unfortunately, however, there is a gap between his definition and his usage of the term "regime," as well as a difficulty in his conceptualization of the Philippine state.

Hawes introduces the concept of "regime" from Fernando Cardoso's work on authoritarian regimes of Latin America (p. 49). Cardoso's definition is concerned primarily with institutions:

> By "regime" I mean the formal rules that link the main political institutions (legislature to the executive, executive to the judiciary, and party system to them all), as well as the issue of the political nature of the ties between citizens and rulers (democratic, oligarchic, totalitarian, or whatever).[79]

[76] Indeed, when in 1981 external funds became harder to come by, it was useful for Marcos to promote chief technocrat Virata. Wurfel should not take this as an indication that the technocrats were temporarily edging out the cronies (p. 254). The same year a more ideological technocrat, Economic Planning Minister Gerardo Sicat, was "fired for raising objections to the costly projects of the First Lady"; see Fabella (fn. 52), 197.

[77] Interview, Francisco S. Tatad, former information minister (under President Marcos), August 22, 1989. In Weber's words, "The ruler's personal discretion delimits the jurisdiction of his officials"; Weber (fn. 3), 1029.

[78] Weber (fn. 3), 1098; see also p. 1028.

[79] Cardoso, "On the Characterization of Authoritarian Regimes in Latin America," in David Collier, ed., *The New Authoritarianism in Latin America* (Princeton: Princeton University Press, 1979), 38.

Based on this definition, if one asks what kind of regime it is, the answer might be "an authoritarian regime." In a patrimonial system, however, where "the ruler's purely *personal ability* to assert his will is to a very high degree decisive,"[80] such an answer may be correct but insufficient. In the case at hand, it is important to specify as well that we are discussing the *Marcos* regime. This is in fact how Hawes uses the term, even if it does go beyond Cardoso's more institutionalist definition. When he writes of "personalized rule" and remarks that "Marcos was hardly the typical 'modern' authoritarian ruler," Hawes is speaking of a regime that has at least as much in common with Latin American caudillos as with the *bureaucratic*-authoritarian regimes that Cardoso is discussing.[81] Perhaps the most appropriate, albeit succinct, definition of the term comes from Wurfel: " 'Regime' refers to both institutions and those who wield power within them" (p. 75 n. 1).

In his discussion of the coconut industry, Hawes concludes that "the nower of the state was used for individual political goals" (p. 82). In his :cond case study, on the sugar industry, he shows that "the vast bulk of ıe surplus went . . . to the personal and political needs of the First Famy" (p. 128) and that Marcos was able to "break the power of the preıartial law sugar bloc" (p. 98). The role of *individual* interests is clear: ne way in which Marcos weakened old elites "was to create monopolies ɔ control the export of coconuts and sugar, place these monopolies ırmly under the control of presidential friends, and use the monopolies s a vehicle to accumulate surplus. . . . [They were] a foundation for his egime" (p. 127).

It is therefore essential to note both the personalistic and the instituional aspects of the regime. As Wurfel explains in his most extensive reatment of the subject of patrimonialism, the authoritarian nature of he Marcos regime lies *between* comparative cases of more thoroughly ıureaucratic and more thoroughly personalistic authoritarianism:

> The Philippine system was too complex and too nearly modern to be called simply "patrimonial." Nor was its power based primarily on an effective

[80] Weber (fn. 3), 1042; emphasis in original.

[81] While caudillos imposed personal orders by decree, Cardoso explains, most modern miltary regimes have a more bureaucratic, institutional basis of power; Cardoso (fn. 79), 35–36. The polity and economy from which the Marcos regime emerged is larger and seemingly nore institutionalized and complex that that of most so-called "caudillos" of modern Latin America. Nevertheless, it is probably easier to draw comparisons with personalistic caudillo egimes (such as that of Somoza) than with the bureaucratic-authoritatian regimes of the outhern Cone (Brazil, Argentina, Chile, and Uruguay), where the military ruled *as an institution*. It would be a valuable project to explore further the similarities and differences etween Marcos and the modern caudillos.

> bureaucracy. Thus the most accurate label for the martial law regime was "patrimonial authoritarianism," or perhaps "neopatrimonial"—highly personalized for its size but more institutionalized than an Ethiopian monarchy or a Dominican or Nicaraguan dictatorship. (p. 153)[82]

It is hard to reconcile Hawes's excellent data on agricultural expo industries, which point clearly to regime accumulation, and his quit abstract theoretical discussion, which posits the "relative autonomy" (the state. He writes that "in addition to being powerful and autonomou the state in the Philippines has indeed been an arena for class conflict (p. 53). Another very general assertion is that "the state is a class stat defending the interests of the bourgeoisie as a whole; however, the stat is relatively autonomous in the sense that it can take and has taken stron state action detrimental to certain segments of the bourgeoisie" (p. 136 Hawes's example of such a "relatively autonomous" action is the state' "attack on the ISI bourgeoisie" (p. 141), which in fact never occurred (a discussed above). But when he taps his own data, he gives clear-cut dem onstration of the state's "coercive and administrative powers" being use for *private* ends (p. 82, also p. 153).

Given the weak separation of the "private" and "official" spheres, i becomes very problematic to assert the "strengthening of the Philippin state" and its "insulation from the demands of civil society," when in fac those who were supposedly insulated (the technocrats) "had no power t resist the demands of Marcos and his cronies" (p. 144). It is more accurat to understand the plunderous activity of Marcos and his cronies as espe cially powerful "demands of civil society." The Marcos regime embodie potent societal interests that seized control of the state and proceeded t centralize access to that state. There was heightened state *capacity* through bureaucratic reform and the use of technocratic expertise, but those technocrats themselves were not *autonomous* in relation to regime interests.[83] Given the overall strength of "the demands of Marcos and his cronies," it is difficult to understand how there could have been signifi-

[82] If one were to try to determine precisely *where* the Philippine case is situated amid more extreme cases of personalistic and bureaucratic authoritarianism, one would need to conduct careful comparative analysis of specific arenas of state policy. Within each arena, it would be necessary to determine whether there was any significant change over time in the degree of separation of "official" and "private" spheres. Obviously, such an effort is beyond the scope of this article.

[83] State autonomy is defined above (fn. 28). State capacity, Kathryn Sikkink writes, "involves the administrative and coercive abilities of the state apparatus to implement official goals"; Sikkink, "State Autonomy and Developmentalist Policy in Argentina and Brazil: Frondizi and Kubitschek Administrations Compared" (Paper presented at the Latin American Studies Association Congress, 1988), 3, quoted with author's permission). In a system with clear-cut patrimonial elements, this definition needs to include not only "official goals," but also the personal goals of the political leadership.

;ant "strengthening of the Philippine state." Other societal interests were ndeed excluded, but that does not point to the existence of a "relatively ıutonomous" state.[84]

Hawes's major difficulty stems from introducing notions from a *bu-eaucratic* state into his analysis of a state that exhibits enduring *patri-nonial* characteristics. His discussion of the state, at points, seems to

> presuppose *the conceptual separation of the "state,"* as an abstract bearer of sovereign prerogatives and the creator of legal norms, *from all personal authority of individuals*. These conceptual distinctions are necessarily remote from the nature of pre-bureaucratic, especially from patrimonial and feudal, structures of authority.[85]

Hawes's empirical data, however, lead him to downplay this "conceptual eparation." His emphasis on *regime* use of *state* power (as well as his use of the term "quasi state monopolies") highlights how remote this concep-ual distinction often is from Philippine reality (p. 128).

Although Hawes does not acknowledge it, the sugar and coconut rading monopolies provide excellent examples of the patrimonial nature f the Philippine state. Weber asserts that under patrimonialism, "only ertain types of capitalism are able to develop ... [and] these forms ... ften reach a very high level of development." Among these types are politically oriented capitalism" and "capitalist wholesale trade"; trade nonopolies, he writes, are especially important in the development of entralized patrimonialism.[86]

With the establishment of the coconut levy in 1971, government levies vere turned over to "private organizations that claim[ed] to represent .. the entire industry" (Hawes, 69). After martial law, the levy became nore onerous, and Marcos "favorites," notably Cojuangco and Enrile, ffectively took control of the levy proceeds (and, by the end of the de-

[84] A largely parallel line of analysis is found in Haggard's insightful discussion of the ature of "weak authoritarian regimes"; he states that "technocrats may gain autonomy from nterest group and legislative pressures under authoritarian rule, but they do not necessarily ain independence from the executive." This executive has a "*personal stake* ... in the crony nterprise"; Haggard (fn. 48), 217–18.

Of the works discussed in this article, Hawes's book alone devotes significant attention to tate theory. But Wurfel, too, in his occasional mention of the subject, asserts that in the early ears of martial law the state had "autonomy from the dominant class" (p. 333) and had the capability to make and implement policy" in a way that was unresponsive "to interests utside the state structure" (p. 334). Only later did the "state [serve] the interests of the uperpatron and his closest clients—the essence of neopatrimonialism" (p. 333). This "essence f neopatrimonialism," I am arguing, was present from the beginning.

[85] Weber (fn. 3), 998; emphasis added.

[86] Weber (fn. 3), 1:240, 3:1091 and 1092. If Weber is the guide, this effectively refutes Montes's (fn. 26) assertion that "a 'rent-seeking' economy cannot readily be classified as cap-alist" (p. 65). In Montes's analysis (fn. 31) the Philippines suffers from a "feudal social basis f political power" (p. 135).

cade, of the milling and export of most coconut products) in orde
build a diversified agricultural, financial, and manufacturing empire.
1983 their United Coconut Planters Bank had assumed control of
country's largest firm, San Miguel Corporation. "Official" powers of r
enue collection were invested in "private" hands, and the result is a cl
indication of what enormous opportunities for patrimonial accumulati
can come from long-term, unchallenged access to the Philippine state.

THE RESURGENCE OF THE "OLD OLIGARCHY"

The declaration of martial law signaled Marcos's attempt to move fro
a decentralized to a centralized patrimonial polity. The most color
description of this process comes from Anderson:

> From one point of view, Don Ferdinand can be seen as the Master Cacique or Master Warlord, in that he pushed the destructive logic of the old order to its natural conclusion. In place of dozens of privatized "security guards," a single privatized National Constabulary; in place of personal armies, a personal Army; instead of pliable local judges, a client Supreme Court; instead of myriad pocket and rotten boroughs, a pocket or rotten country, managed by cronies, hitmen, and flunkies.[87]

Marcos shared the patrimonial ruler's dislike of "independent ec
nomic and social powers." He faced "a property-owning, especially .
land-owning stratum of . . . subjects [that had] easily monopolize[d] t
local offices,"[88] and he was determined to undercut those wel
entrenched oligarchic forces that might stand in the way of his centra
izing agenda. Yet at the same time, the very fact that they controll
crucial independent resources meant that he could not afford to antag
nize them as a group. In a world of mobile capital, Marcos knew that l
could not "induce business performance" if he acted too generally an
too rashly against his rivals.[89] As mentioned above, he had to move sele
tively against those who most threatened his regime.

Weber explains the classic conflict of "patrimonial ruler versus loc
lords":

> The continuous struggle of the central power with the various centrifugal local powers creates a specific problem for patrimonialism when the pat-

[87] Anderson (fn. 10), 20. [88] Weber (fn. 3), 1040.

[89] The idea of inducement comes from Charles Lindblom's discussion of "the privileg position of business." Although he focuses on advanced capitalist countries, there are non theless parallels to be found in Marcos's situation after 1972. Especially in the beginning th ruler's household controlled but a portion of the national economy, and Marcos needed t ensure (domestic and foreign) investors' confidence in general business conditions. See Lind blom, *Politics and Markets: The World's Political-Economic System* (New York: Basic Book 1977), 170–88, quote from p. 174.

rimonial ruler, with his personal power resources—his landed property, other sources of revenue and personally loyal officials and soldiers—confronts not a mere mass of subjects . . . but when he stands as one landlord . . . above other landlords, who . . . wield an autonomous authority of their own. . . . *The patrimonial ruler cannot always dare to destroy these autonomous local patrimonial powers.*[90]

As we have seen, Marcos's own "personal power resources" were greatly augmented by international capital inflows. Using these resources, he attempted to "centralize rather than disband the national network of patron-client relations—to become 'supreme godfather' " (Wurfel, 152). His immediate dissolution of Congress, of course, had undercut rival patronage networks. And just as Weber notes that "the 'father of the people' . . . is the ideal of the patrimonial states,"[91] so the First Couple attempted to promote themselves as "Father and Mother of the Nation" in barrio-level youth organizations (Aquino, 86).

But Marcos was never able to contemplate the systematic destruction of independent power bases. Nor did he ever show an interest "in upsetting the established social order."[92] While the regime had a major influence on the composition of the country's elite, it ultimately faced a full-scale resurgence of rival centers of power. As we shall see, opposition from these forces began to build in the early 1980s, and in 1986 they contributed in a major way to his downfall.

A closer analysis of exactly how Marcos moved against the Lopez clan in 1972 highlights his concern not to antagonize capital holders *in general*. The expropriations of the Lopez newspaper and television studios were "crudely" accomplished, and the facilities soon found their way into the hands of Benjamin Romualdez and Benedicto. But in the case of the Manila electric company, the Lopez's Meralco, " 'Marcos knew that it would be scandalous to use the same crude methods.' "[93] It was a corporation with twelve thousand stockholders, some of them foreign creditors. Marcos effected this expropriation by arresting one of the Lopez sons, and "with an effective hostage to force compliance, Marcos used the full resources of the state to break the Lopez hold over Meralco step-by-step." By late 1973, utilizing such techniques as cuts in power rates and withholding of credit, "Marcos gained control of U.S. $5.7 million in Lopez assets for only $1,500." The holdings of smaller stockholders, however, were not absorbed by the regime.[94]

[90] Weber (fn. 3), 1055; emphasis added.
[91] Ibid., 1107.
[92] Anderson (fn. 10), 22.
[93] McCoy (fn. 30), MS p. 45, quoting an affidavit of Oscar Lopez.
[94] Ibid., 45–48.

Similarly, landowners tended to be punished only to the extent that they represented a threat to the regime. As Wurfel explains:

> Land reform's most important political function was to strike a blow at the "oligarchy," those wealthy elites who formed the core of his political opposition. The Aquino estates were among the first to be expropriated. The president lost his originally keen interest after owners of more than one hundred hectares had been dispossessed. (p. 166)

Moreover, by excluding sugar and other export crops from land reform, Marcos was being "cautious not to antagonize all landed wealth at once" (p. 167).

The Marcos regime also trod softly in the area of fiscal reform in order to avoid antagonizing powerful business interests. Authoritarian government was unable to overcome the state's historical inability to tax the monied classes: real property and personal income tax rates remained low, and—just as in the pre-martial law years—indirect taxes continued to supply 70–80 percent of total tax revenue for the state. The "new government could codify and simplify tax laws and administration" and grant tax and tariff exemptions to its allies, but "it did not have enough power to threaten the economic power of the groups that had been dominant in the pre-martial-law period."[95] And, as discussed above, government policies did little to undermine the trade barriers behind which ISI entrepreneurs thrived.

Until roughly 1981 much of the pre-martial law elite had little to complain about. Besides those who were excluded outright, a second gruop of the pre-martial law elite families "managed to hold their own," and a third, "favored" group expanded its business interests considerably.[96] There must have been resentment at the First Couple's reported "squeeze" of businesspeople (Wurfel, 137), which was apparently most rampant against the more politically vulnerable Filipino-Chinese business community. And when nonfavored firms went bust, the regime seemed to have a "kick 'em when they're down" approach to corporate acquisition: in the course of the 1970s seven banks—all in a weak position—ended up in crony hands. But these actions were largely selective. In terms of overall business confidence, high commodity prices and external funds promoted respectable growth rates through the 1970s.

[95] Montes (fn. 31), 96, 114–15, 117, 120. Attempts at a more centralized patrimonial polity, then, did not resolve the revenue dilemmas faced in the pre-martial law period, when many oligarchic patrons competed in the electoral arena.

[96] John F. Doherty, "Who Controls the Philippine Economy: Some Need Not Try as Hard as Others," in Belinda A. Aquino, ed., *Cronies and Enemies: The Current Philippine Scene*, Philippine Studies Occasional Paper No. 5 (Honolulu: Philippine Studies Program, Center for Asian and Pacific Studies, University of Hawaii, 1982), 30.

Moreover, whether crony or noncrony, the oligarchy as a whole was no doubt appreciative of how the regime had greatly curbed popular demands.

Significant elite opposition to Marcos did, however, become evident after the financial crisis of 1981, when a prominent Filipino-Chinese businessman fled the country without settling nearly $85 million in debts. Several financial institutions closed down, and credit became very tight. A massive Central Bank bailout of failed crony firms provoked a great deal of resentment, and a suprafamilial elite coalition based on clear differences of economic interest developed among "nonfavored" and "less-favored" businesspeople. The most prominent business organization seeking political change was the Makati Business Club, formed in 1982.[97]

Marcos was, in fact, facing increasing opposition from much of the "old oligarchy" of pre–martial law origins. In their own analysis, crony excesses were largely to blame for constricted economic opportunities. First, the bailout of crony firms contributed heavily to the scarcity of credit. Second, growing international perceptions of unbridled cronyism exacerbated the regime's problems in dealing with increasingly tight global credit markets. Especially after the assassination of Benigno Aquino in 1983 and the capital flight that came in its wake, international creditors lost their faith in Marcos; particularly damaging was the falsification of international reserve figures by the Central Bank. "An important reason the big business sector rejected the Marcos regime was its loss of credibility with international creditors and consequent inability" to furnish the economy with essential external resources.[98] Because manufacturing interests remained heavily dependent on imported inputs, and hence foreign exchange, the economy went into a tailspin. Later, many of the influential members of the Makati Business Club became key backers of the presidential campaign of Corazon Aquino in 1986; significantly, one of her successful tactics was to call "a boycott of Marcos-controlled newspapers, government and crony-owned banks, and the San Miguel commercial empire" (Wurfel, 300).

Many members of this anti-Marcos coalition assumed key posts in the Aquino administration. Once in power, these oligarchs had an opportunity to launch a concerted attack on those aspects of the political economy that had long retarded the development of more advanced forms of capitalism in the Philippines. In Weber's terms, "politically oriented capi-

[97] See de Dios (fn. 60), 102–3. Broad attempts to link part of this elite opposition to the "national entrepreneurial class" (p. 105; see also p. 3). But a close examination of either the "bourgeois bombers," to whom she is referring, or other major groups, especially the influential Makati Business Club, makes this implausible.

[98] De Dios (fn. 5,), 133.

talism" and "capitalist wholesale trade" had flourished under Marcos, but other forms of capitalism had been hindered by "the wide scope for actual arbitrariness and the expression of purely personal whims on the part of the ruler and the members of his administrative staff."[99] As Wurfel points out, cronies in manufacturing ventures were "less stable" than those who were given "control over marketing or processing a major crop" (p. 237). And within the export-crop monopolies, there were often limits to the types of economic activity that were attempted (Hawes, 76, 102–3). There were, indeed, some relatively sophisticated activities taken over by the cronies, including commercial banks and the San Miguel Corporation. But while a few enterprises prospered, crony enterprises on the whole contributed significantly to the country's financial disaster in the early 1980s.[100] State-owned enterprises, which had more than tripled in number between 1972 and 1984, were patronage-ridden beasts that soaked up 30 percent of public expenditures by the late Marcos years.[101] Marcos's attempt at centralized patrimonialism led, in the end, to an unprecedented degree of "unpredictability and inconsistency"[102] in the functioning of the state apparatus, as "personal whims" triumphed.

Did the oligarchs of the Aquino administration, then, attempt to create "the political and procedural *predictability*, indispensable for capitalist development, which is provided by the rational rules of modern bureaucratic administration"?[103] To be sure, there were important efforts at liberalization and privatization, as well as major initiatives to regain the "ill-gotten" wealth of the Marcos cronies. But on the whole, these efforts failed.[104] And more fundamentally, one can still say that political admin-

[99] Weber (fn. 3), 1:239.

[100] See the influential white paper of the University of the Philippines school of economics; Emmanuel S. de Dios, ed., *An Analysis of the Philippine Economic Crisis: A Workshop Report* (Quezon City: University of the Philippines Press, 1984).

[101] Stephan Haggard, "The Philippines: Picking Up after Marcos," in Raymond Vernon, ed., *The Promise of Privatization* (New York: Council on Foreign Relations, 1988), 93, 97. Within these enterprises, the public-private distinction was often blurred beyond recognition; as Haggard reports, prior to a 1985 Supreme Court ruling, "several managers of state-owned enterprises even denied that they were public officials!" (p. 95). See also Weber's discussion (fn. 3), 1097, of the extensive growth of state enterprises in patrimonial polities.

[102] Weber (fn. 3), 1095.

[103] Ibid.; emphasis in original.

[104] ISI monopolists effectively control the policy-making agenda in the Aquino regime, and

> international institutions seeking to push rationalizing or liberalizing strategies such as export-oriented industrialization are in the queer position of not having a clear constituency with political influence. . . . [T]he free-trade ideology . . . is being resisted by the traditional elite (now holding key positions in the government) who have been accustomed to privileged, rent-producing positions in the economy. (de Dios [fn. 5], 140–41)

As early as 1988 Haggard concluded that "barriers to rapid privatization emerged, and implementation has lagged behind state intentions." In light of subsequent experience, this conclusion could now be stated even more forcefully. See Haggard (fn. 101), 92.

Economist Gustav Ranis, a longtime consultant on Philippine economic issues, wrote in

istration is often treated as a personal affair. Patrimonial features endure, as the highly arbitrary personal discretion of oligarchs currently in power largely determines the assignment of privileges granted by the state. Once again, as in the pre–martial law years, a decentralized polity simply gives *more* oligarchs a chance to claw for the booty of state. If Weber is correct, more advanced forms of capitalism will not flourish until the Philippine state achieves some minimal level of success in supplying "continuity, trustworthiness and objectivity of the legal order, and ... the rational, predictable functioning of the legal and administrative agencies."[105] As long as oligarchs and cronies continue to plunder the state, it is difficult to see how such success will be achieved.

Conclusion

"What are we in power for?" responded a pre–martial law senator when reporters asked him about corruption charges (Wurfel, 100). Like many oligarchs and cronies in the Philippine state, this politician knew very well that the weak distinction between "official" and "private" spheres means ample opportunity for plunder of the state. Marcos elevated the practice to its highest level of development, but even in his absence, patrimonial elements remain strong. Cheap credit, tax incentives, state licenses, and monopoly privileges hinge on "personal considerations." State resources are blatantly appropriated for private ends, and a new group of cronies has been spawned. Indeed, certain senators in the Aquino era consider "plunder" such an enduring plague upon the land that they introduced a measure declaring it punishable by death. Persons amassing $2.3 or $4.6 million in "ill-gotten wealth"—the sponsors were in disagreement as to the proper threshold—would be executed by the state they had sacked.[106]

1989: "While the Philippines is certainly not alone in preferring [to avoid] ... painful economic restructuring, it is undoubtedly one of the foremost practitioners of the art"; Ranis, *Far Eastern Economic Review*, September 28, 1989, pp. 153–54, at 154.

Even the prosecution of Marcos cronies has been plagued with failure, and only a tiny portion of the "ill-gotten wealth" has been recovered. See the *Philippine Daily Globe*, December 16, 1989. Quite early on, in fact, the major agency tasked with the effort (the Philippine Commission on Good Government) was already being called a "tainted watchdog" because of alleged misappropriation of seized assets. See the *Far Eastern Economic Review*, September 17, 1987, pp. 22–27.

[105] Weber (fn. 3), 1095. This is not to say, of course, that corruption is incompatible with advanced capitalism: one can think of myriad examples where the two thrive simultaneously. In Weber's analysis bribery and corruption have the "least serious effect" when they are calculable and become most onerous when fees are "highly variable" and "settled from case to case with every individual official" (vol. 1, p. 240). If correct, a major obstacle to the development of more sophisticated forms of capitalist accumulation is not corruption per se, but highly arbitrary corruption.

[106] *Manila Chronicle*, June 11, 1989.

Despite its origins in "people power," the best characterization of the Aquino government is simply "restorationist."[107] The following description of the Philippine political system, written by the Central Intelligence Agency in 1965, remains remarkably accurate. After noting a history of domination by a "small, wealthy elite," it asserts

> the primacy of the kinship group over all institutions, including the state. ... [K]inship and personal connections are far more important than merit or legal niceties in political, social, and business relations; this has contributed to widespread acceptance of nepotism and corruption as the normal road to political and personal advancement. A corollary to the importance of personal relations is the widespread disrespect for the impersonal rule of law.[108]

Marcos's authoritarian government did not bring fundamental change to the particularistic, plunderous way in which dominant interests interact with the Philippine state, and with his departure came a fundamental resurrection of the more decentralized pre–martial law system. Such a high degree of continuity needs to be explained.

The thesis presented here is that this continuity results from ongoing patrimonial features of the Philippine state and that current scholarship on the Marcos regime and the Philippine political economy would benefit greatly from a more careful examination of these features. Belinda Aquino's focus on Marcos the individual is entirely compatible with a patrimonial framework, but it is essential to understand the nature of the state that he and his cronies plundered. "Bonapartist" explanations are problematic in their assertion of "relative autonomy" and in the weak emphasis on regime interests. (Moreover, it is difficult to show the presence of a "pliable mass organization" of any significance.) Broad relies excessively on a "transnationalist-nationalist" distinction that is unable to account successfully for cronyism and its persistence. Her assertion of a "reshaping of the entire Philippine political economy" would not lead one to expect any significant continuity across regimes. Hawes brings regime interests into his analysis but often crowds them out in his efforts to define which elite segments (each with supposed coherence of economic interest) control the state. His useful insights on regime use of state power are often obfuscated by abstract theorization.

Of all the works reviewed, Wurfel's puts the greatest stress on continuity across regimes (pp. 74, 274, 323, 326, 336), and his book alone integrates patrimonial terms. Wurfel's theoretical framework is quite var-

[107] De Dios (fn. 5), 124.

[108] CIA, *Philippines: General Survey*, National Intelligence Survey, NIS 99, July 1965 (sanitized copy released November 1980), excerpted in Daniel B. Schirmer and Stephen Rosskamm Shalom, eds., *The Philippines Reader: A History of Colonialism, Neocolonialism, Dictatorship, and Resistance* (Boston: South End Press, 1987), 126.

ied, however, and patrimonialism is only one of many themes found throughout the book. In some cases, accounts of patrimonial activity (for example, the extensive use of Marcos relatives as emissaries, candidates, and regime spokespersons) are not explicitly linked to an overall analysis of the strength of patrimonial patterns. At one point, the pre–martial law system seems to be given patrimonial attributes (p. 325), but in general he confines the term to the Marcos years. In short, patrimonial elements within the Philippine state deserve the same kind of careful examination that Crouch has given to the patrimonial elements in Indonesia and that Thomas Callaghy has given to the "patrimonial administrative state" in Zaire.[109]

It would seem that the most important task of future research should be to explain *why* patrimonial features have persisted in the Philippines despite enormous change (for example, in regime type, in increasing levels of economic complexity and "modernization") and yet elsewhere seem to have subsided in the face of change. "Patrimonialism" is a label under which one can find enormous variation according to the relative historical strengths of state and society, the nature of domestic coalitions, and the economic and strategic relationships of a country to external forces. Closer examination of these factors will be useful both in developing a clearer and more comprehensive model of "comparative patrimonialism" and in explaining the *persistence* and *evolution* of patrimonial features. The "bureaucratic polity" in Thailand, for example, displayed many patrimonial characteristics, but the nature of that patrimonialism—in which a bureaucratic elite extracted privilege from a historically weak business class—contrasts in many ways with patrimonial features in the Philippines, where a powerful oligarchic business class extracts privilege from a largely incoherent bureaucracy. Likewise, the current evolution of relations between the state and the business community in Thailand contrasts with the basic continuities that have endured in the Philippine setting.

In the midst of enormous change there has been no effective pressure from either external or domestic forces to undermine the patrimonial features of the Philippine state. For largely strategic reasons, of course, the system has been continually funded by external forces.[110] And be-

[109] See Crouch (fn. 11); and Callaghy (fn. 11).

[110] As Gustave Ranis wrote in 1989, discussing the country's strong standing with the U.S. and Japan,

> There is likely to be overwhelmingly strong "need to lend" for geo-political reasons, on the one hand, and for the "need to borrow" to avoid domestic adjustment, on the other. As a consequence the system will probably find it possible to continue its non-optimal growth path for some time to come.... [T]he very economic opportunity that the Phil-

cause patrimonial features in the Philippines reward oligarchs accordir to their proximity to the political machinery, those who are temporari "on the outs" exert far more effort in trying to restore such proximi than in pressing for profound structural change.

This article has highlighted patrimonial features found in the relatioı ship between state and dominant economic interests, but the frameworı would also have a high degree of utility in discussing the Philippine mi itary (see Wurfel, 256), as well as local politics[111] and other areas. It also important to emphasize that a focus on oligarchs and cronies in tł Philippine state by no means implies that analyses of the state's relatioı with *domestic popular forces* or *external actors* are any less important i developing a comprehensive understanding of the Philippine state. I discussing "patterns of class domination" (p. 162), for example, Hawes entirely correct to show which side the state is on, and it is important t emphasize the crucial role played by the U.S.-supported system of stat repression in protecting plunderers from revolutionary upheaval.[112]

Building on Broad's work, a fuller understanding of the Philippin state requires clear analysis of why external actors remain so powerful i Philippine policy formulation, when other economies in the region th used to be similarly "dependent" have since managed to carve ou stronger roles for themselves. For a start, it is worth inquiring as to hov cliental relations with the United States seem to have insulated Philip pine elites from any real sense of intrastate competition—competitio that has often been the historic starting point for serious state-buildin projects.[113] For most of this century the oligarchy's major external con cern has been how to ensure continued U.S. sponsorship for thei domestic hegemony. The plunder of the Philippine state is not sel sustaining; ultimately, it depends on the international dole. As U.S. stra tegic perceptions change, and the oligarchy finds it increasingly difficu to use the military bases to extract critical support from Washington pa trons, one might begin to see changes as well in the nature of the Phil ippine state.

ippines' geo-political importance provides threatens to be the undoing of its political economy. (Ranis [fn. 104], 154)

[111] As attested by the current research of John Sidel of Cornell University; see Sidel, "Bi Men with Bolos, Bosses, Bullets, Bank Loans, and Bus Companies: Local Power in Philippin Politics" (Unpublished manuscript, 1990).

[112] When it comes to arrests of labor leaders or military harassment of peasant communi ties, of course, the state is very clearly an "independent actor" vis-à-vis societal forces.

[113] Joel Migdal, for example, writes that "a prime motivation for state leaders to attemp to stretch the state's rule-making domain within its formal boundaries, even with all th risks that has entailed, has been to build sufficient clout to survive the dangers posed by thos outside its boundaries, from the world of states"; Migdal, *Strong Societies and Weak State* (Princeton: Princeton University Press, 1988), 21.

DOES MODERNIZATION BREED ETHNIC POLITICAL CONFLICT?

By SAUL NEWMAN

Fouad Ajami, *The Vanished Imam: Musa al Sadr and the Shia of Lebanon*. Ithaca, N.Y.: Cornell University Press, 1986, 228 pp.
Donald L. Horowitz, *Ethnic Groups in Conflict*. Berkeley: University of California Press, 1985, 697 pp.
Joseph Rothschild, *Ethnopolitics: A Conceptual Framework*. New York: Columbia University Press, 1981, 290 pp.
Anthony D. Smith, *The Ethnic Revival*. Cambridge: Cambridge University Press, 1981, 240 pp.
Edward A. Tiryakian and Ronald Rogowski, eds., *New Nationalisms of the Developed West: Toward Explanation*. Boston: Allen and Unwin, 1985, 394 pp.

IN the last decade social scientists devoted considerable attention to the study of ethnic political activity throughout the world, a renewed interest that came in response to empirical and theoretical circumstances. Since the early 1960s there have been many manifestations of ethnic political conflict in both industrialized and developing states. This conflict has resulted in profound political change and personal suffering in many of the well over one hundred nation-states that contain significant ethnic minorities. The theoretical tools available to social scientists in the early 1960s were unable to account for the persistence of these conflicts. As a result, scholars have more recently undertaken to study anew various aspects of ethnic political activity, in the hope that theory can catch up to empirical reality. Despite these efforts, however, considerable gaps still remain in the theoretical literature. Much of the study of ethnic politics since World War II has been conditioned by the study of economic modernization and political development as delineated in the writings of Karl Marx and Emile Durkheim. Because the ethnic politics literature evolved within the confines of these modernization and development paradigms, it necessarily tended to emphasize certain issues and failed to address others.[1] Until the early 1970s scholars argued that ethnic conflict would subside with the emergence of "modern" societies. When this failed to occur, these "melting pot modernization" theories were re-

[1] Although there were ethnic politics theories that did not come under the modernization rubric, they were certainly in the minority. In this review article I will focus on this dominant strain within the literature.

placed, in the mid-1970s, by "conflictual modernization" theories, whic argued that growing interethnic social and economic activity would in crease the likelihood of interethnic conflict. Economic modernizatio was thereby elevated to the level of a sufficient condition for the emer gence of ethnic political conflict. Emphasis was placed on the social an economic changes that preceded the emergence of ethnic conflict and o the political elites that exploited these changes, but little attention wa paid to the structure and composition of ethnic political movements o to the content of ethnic political debate. Since the mid-1980s proponent of a "psychological" approach to the study of ethnic political conflict have tried to develop a new theoretical framework that would redress many of the weaknesses of the conflictual modernization approach. They have been unable to articulate a logically consistent alternative, however.

The weaknesses of the ethnic politics literature stem from a misinterpretation of the relationship between modernization and ethnic conflict. Because of their dependence on the modernization approach, conflictual modernization theorists, such as Walker Connor and Anthony Smith, have predictably emphasized certain issues. An examination of the evolution of the ethnic politics literature will shed some light on the manner in which social science paradigms define, and sometimes constrain, the investigation of political phenomena. All too often, important areas of investigation in the study of ethnic politics have been ignored because of paradigmatic closure. But in the 1990s, with the world's attention focused on ethnic political conflict in Canada, Eastern Europe, and the Soviet Union, it is particularly important that researchers throw off these constraints.

The problem is that students of ethnic politics have failed to exploit the theoretical implications of the modernization process to its fullest. They now need to develop a new ethnic modernization theory that highlights the relationship between modernization and the institutionalization, composition, and ideologies of ethnic political movements. In the course of modernization ethnic political identities and institutions are repeatedly created and re-created anew, a process that constantly destroys the old ethnic loyalties while creating new ethnic ideologies, institutions, and constituencies. The process of modernization explains not only the origins of ethnic conflict but also the form of that conflict, and the success or failure of specific ethnic political movements.

We will attempt to reinvigorate the conflictual modernization approach and test its utility by briefly examining cases of ethnic political conflict that have emerged in Western Europe and North America since the late 1960s. As we shall see, the success of ethnoregional political par-

ties such as the Parti Québecois (PQ) in Quebec, the Scottish National Party (SNP) in Scotland, and the Volksunie (VU), Rassemblement Wallon (RW), and Front Démocratique des Francophones (FDF) in Belgium can be attributed to the process of social and economic modernization that accompanies the emergence of a postindustrial society.[2] These cases illustrate that conflictual modernization theory can do more than identify the variables that give rise to ethnic political movements. It can also be used to explain and analyze the ideologies and organizations of ethnic political movements in both developing and industrialized states.

By examining how the modernization process provides the ideological and organizational resources for new ethnic movements, we can begin to merge the study of ethnic conflict with the study of new social movements and resource mobilization. From this endeavor may emerge a comprehensive explanation of why and how certain ethnic identities become politicized and why some ethnic movements succeed whereas others fail.

Ethnicity: From "Primordial" Obligation to Modern Identification

Twentieth-century approaches to the study of ethnicity in politics can be traced back to the writings of Marx and Durkheim, both of whom evaluated ethnic identities as part of a larger set of phenomena subject to transformation by the forces of economic modernization. For Marx, ethnic nationalism and religion were both integral parts of the superstructure created by the dominant economic and political classes.[3] Like religion, ethnicity could be used by the dominant classes to legitimize their rule, bind together different classes through the creation of a false consciousness, and foster political passivity among the economically weaker classes. The "idea" of the ethnic group was transformed into a "higher idea" than the reality that flowed from the material conditions of existence. The process of modernization was expected to culminate in the victory of the proletariat, meaning that ethnic nationalism would ultimately disappear as would other social manifestations of class domination. The Marxist scholarly legacy, which influenced Marxists and non-Marxists alike, was to analyze the role of ethnicity in politics as a

[2] For a detailed discussion of these cases, see Saul Newman, "The Ethnic Dilemma: The Rise and Decline of Ethnoregional Political Parties in Scotland, Belgium, and Quebec" (Ph.D. diss., Princeton University, 1989).

[3] Marx writes about the larger issue of consciousness flowing from the material conditions of existence; see Marx, "The German Ideology: Part I," in Robert C. Tucker, ed., *The Marx-Engels Reader*, 2d ed. (New York: W. W. Norton, 1978), 158, 159, 164, 165.

comparatively ephemeral phenomenon, to be shaped, and eventuall destroyed, by the forces of modernization.

Durkheim argued that the process of economic modernization lead to a division of labor,[4] which has the potential to replace a mechanicall integrated society with an organically integrated society. A mechanicall integrated society is united by a collective conscience created by a serie of "primordial" identifications. These remarkably enduring primordia identifications contain the sacred symbols that reproduce the social struc ture of a society. Consequently, the collective conscience plays an indis pensable role in the maintenance of social unity. With the change to a society based on the division of labor, however, every citizen becomes dependent on every other citizen, because no one person can be self-sufficient. Each person is a small piece in a huge puzzle that can only be completed when each performs his or her particular role. In this organically integrated society—the product of economic modernization—the primordial collective conscience is no longer relevant to the integration and survival of society. Ethnic identification, having been rendered dysfunctional, will therefore disappear.[5]

Although many scholars endorsed either Marx's or Durkheim's melting pot modernization approach, the weight of empirical evidence eventually overwhelmed these theoretical arguments. It became increasingly obvious in the 1960s that there was something fundamentally wrong with their dominant paradigms as they related to ethnic political conflict.[6] As such conflicts increased in number and scope, they were perceived as more than just a stubborn relic of a bygone era. Rather, ethnic conflict in the Third World was considered a result of new social strains.[7] This strain theory was borrowed from the rediscovery of Durkheim's anomie. According to Durkheim, the anomic individual will resort to suicide during the disorienting process of modernization, that is, in the period after the disappearance of the mechanically integrated society and

[4] Durkheim discusses the relationship between modernization and ethnicity; see Durkheim, *The Division of Labor in Society*, trans. George Simpson (New York: Free Press, 1933), chap. 6.

[5] There are two caveats to this statement. Later on we will discuss one of them: the survival of dysfunctional identifications during the process of achieving a mechanically integrated society. The second relates to the possible tension between Durkheim's early writing in *The Division of Labor in Society* and his later *The Elementary Forms of the Religious Life*. In this second work Durkheim is less willing to predict the disappearance of a collective conscience. Nevertheless, the historical development of the ethnic politics literature has been largely dependent on Durkheim's earlier conclusions in *The Division of Labor in Society*.

[6] The mainstream of American political science also began to reevaluate the modernization perspective as it related to the economic modernization and political development of the Third World.

[7] The term "strain theory" comes from the writings of Clifford Geertz. See David E. Apter, ed., *Ideology and Discontent* (New York: Free Press, 1964).

before the creation of the organically integrated society. Strain theorists argued that as ascriptive ties lose their political relevance, unintegrated citizens, looking for an anchor in a sea of changes, will grab hold of an increasingly anachronistic ethnic identity, which bursts onto the scene and then recedes as the process of structural differentiation moves toward a reintegrated society.[8] Thus, ethnicity might resurge temporarily, but like suicide, it is a manifestation of anomie that will inexorably disappear.

This perspective, too, fell prey to the overwhelming weight of evidence against it. Not only did ethnic political conflict continue unabated in the Third World, but it also appeared in the industrialized world. In Northern Ireland the "troubles" resulted in violence of the type once believed to be reserved for Third World states. In Quebec, Scotland, Wales, and Belgium ethnic political movements challenged the legitimacy of centralized states. These challenges were clearly more than just a hiccup in the modernization process, a fleeting disjuncture between differentiation and integration. Proponents of melting pot modernization were forced to rethink its relevance to the study of ethnic political conflict.

Rather than abandoning an analysis of the modernization process, scholars sought to rework the theories dealing with the influence of modernization on ethnic political relations.[9] In a series of articles Walker Connor argued that the process of economic modernization does not undermine ethnic divisions but invigorates them[10] by bringing together previously isolated ethnic groups that suddenly find themselves competing for the same economic niches. Therefore, modernization, rather than resulting in a new form of integration, increases ethnic group interactions that may heighten conflicts. Proponents of this conflictual modernization approach argued that during the process of urbanization people compete for occupational positions and seek out their ethnic identity—

[8] This version of strain theory, prevalent in the 1960s, was delineated in Neil J. Smelser, "Mechanisms of Change and Adjustment to Change,"in Bert F. Hoselitz and Wilbert E. Moore, eds., *Industrialization and Society* (The Hague: UNESCO and Mouton, 1963), 41.

[9] Melting pot modernization theorists such as Geertz had hinted at the role that modernization might play in the enhancement of ethnic conflict. Nevertheless, it was up to other scholars to state this explicitly; Geertz, ed., *Old Societies and New States: The Quest for Modernity in Asia and Africa* (New York: Free Press, 1963).

Some people began to develop newer interpretations of ethnic political conflict that did not depend on the modernization paradigm. Some focused on the role of business or labor rivalries in determining ethnic conflict, while others developed a cultural pluralist model that emphasized the isolation of ethnic groups in maintaining ethnic peace. Yet the dominant interpretation of ethnic political activity remained firmly rooted in the modernization perspective.

[10] A key article in this series is Connor, "The Politics of Ethnonationalism," *Journal of International Affairs* 27 (January 1973), 1–21.

that is, tie their condition to the plight of a larger social group—as a to in promoting their economic and political demands. Of course, the pr cess of modernization could also unite previously separate ethnic grou if their short-term and long-term economic and political interests coi cided.

This reinterpretation of the relationship between modernization an ethnicity had a profound influence on the study of ethnic political con flict. No longer was it dismissed as a primordial identity that would soo be outgrown. Ethnicity was instead seen as an identity that could b constantly created and re-created to suit particular political goals,[11] as tool in the struggle to achieve political goals rather than simply as a com forting environment for the anomic individual. Ethnicity was not a "pri mordial attachment that stems from the 'givens' . . . of social existence."[12]

Crawford Young clarified the new understanding of ethnicity as an identity that could be created and re-created anew to suit particular economic and political circumstances:

> Cultural identities are not usefully described as primordial, but are for the most part relatively new, and vastly expanded in scale. . . . Identity, at bottom is a subjective self-concept or social role; it is often variable, overlapping and situational. . . . "We" is defined in part by "they"; the relevant other in a social setting is central in shaping role selection.[13]

This fungible interpretation of ethnicity had important ramifications. The careful definition of the characteristics of ethnic identity—cultural, religious, territorial, racial, and ancestral markers—became increasingly irrelevant because ethnic identification was viewed as changing to suit changing ends. The scholarly effort turned instead to determining how the process of modernization shaped those forces that resulted in the renewal of ethnic conflict—to addressing the question: What forces promoted ethnic political conflict?[14]

Explaining the Ethnic "Revival": Building on the Conflictual Modernization Approach

In *The Ethnic Revival*, Anthony Smith applies Connor's approach in accounting for a centuries-old tradition of ethnic political activity. Histor-

[11] See Nathan Glazer and Daniel P. Moynihan, *Ethnicity: Theory and Experience* (Cambridge: Harvard University Press, 1975), chap. 1.

[12] Geertz, "The Integrative Revolution: Primordial Sentiments and Civil Politics in the New States," in Geertz (fn. 9).

[13] Young, *The Politics of Cultural Pluralism* (Madison: University of Wisconsin Press, 1976), 65.

[14] A whole new set of theories arose out of this reformulation. Arend Lijphart lists eight

ically, ethnic ties have waxed and waned in importance in response to interstate conflict, sometimes serving as a mobilizing mechanism for war. With the birth of the modern state, tensions between economic modernization and political development fed ethnic conflict. The new bureaucratic state sought its legitimacy in scientific rationality rather than in divine right. Simultaneously economic and educational modernization produced an ever-expanding population of secular intellectuals eager to reinforce this legitimacy and an intelligentsia prepared to staff the expanding bureaucracies. However, when potential bureaucrats moved to the metropolis to take up positions in the state commensurate with their training and education, they found that there was no room for them in the state bureaucracy. Disgruntled, they returned to their ethnic group to lead ethnic political movements that provided an institutional legitimacy for their perceptions of injustice.

Smith turns Weber and Durkheim on their heads by arguing that the processes of economic modernization and political development often act at cross-purposes. By focusing on the process of political rationalization rather than on Weber's ideal types, he is also able to identify the "irrational" results of the rationalization process. However, Smith's parsimonious argument leaves too much unexplained. He does not define ethnicity, viewing that task as a function of the manipulative ideologies of ethnic elites. Moreover, although he can describe a proclivity for ethnic political activity, he cannot completely explain any particular ethnic political revival. For example, to account for the recent rise of ethnic political activity in the Western world, Smith refers to the role played by the collapse of colonial empires, the disjuncture between state boundaries and ethnic groups, and the rise of technocratic managerical politics (chap. 9).[15] He says that declining respect for the state at a time when the state is becoming increasingly centralized and intrusive has led to the mobilization of ethnic movements that attempt to make states smaller and more representative of ethnic boundaries. None of these explanations derives from the conflictual modernization perspective that he claims serves as his intellectual nourishment. The decline of empire and the delineation of national borders have little to do with the effect of modernization on ethnic politics. In addition, the growth of the technocratic state after World

approaches to explain the existence of ethnic political activity in the Western world. See Lijphart, "Political Theories and the Explanation of Ethnic Conflict in the Western World: Falsified Predictions and Plausible Postdictions," in Milton J. Esman, ed., *Ethnic Conflict in the Western World* (Ithaca, N.Y.: Cornell University Press, 1977), chap. 2.

[15] Such explanations might logically have been derived from the rise of ethnic nationalism in the United Kingdom and Belgium. Both these states saw a rise in ethnic conflict as their empires dissolved.

War II should if anything have increased the opportunities for techn cratic managers. Smith's ad hoc explanations, which do not focus on a overproduction of an intelligentsia, lead to a serious disjuncture betwee the necessary and sufficient conditions for the revival of ethnic politic activity.

Furthermore, by focusing on modernization rather than on ethni identities, Smith fails to solidify the connection between the two. Wh should secularizing intellectuals turn to ethnic historicism instead of rad ical secular ideologies such as Marxism? Smith's failure to account ade quately for this decision typifies the failure of the ethnic modernizatio literature to find its ethnic component.

Finally, Smith is content to understand the motivations of elites alone But this leaves empirical gaps in his understanding of ethnic movement because the overwhelming majority of ethnic political movement members are not dissatisfied bureaucratic intelligentsia.[16] They also make up a very small percentage of the membership of ethnic political movements. How, then, do these elites mobilize mass support? More importantly, why should the mass of supporters promote an agenda designed to suit the personal needs of a narrow sector of the intelligentsia? The strengths and weaknesses of Smith's argument were reproduced in the work of many other scholars loyal to Connor's modernization approach.[17] In response to the weaknesses, scholars began to rework the conflictual modernization paradigm again.

Reworking the Conflictual Modernization Paradigm

By the early 1980s some scholars had begun to realize that the popular theories for explaining ethnic political conflict were incomplete. Too

[16] Survey data from Scotland and Belgium indicate that the bureaucratic intelligentsia is actually underrepresented in ethnic political parties compared with the population at large. See Ivor Crewe and Bo Sarlvik, Scottish Election Study, October 1974, available through the Inter-University Consortium for Political and Social Research; and N. Delruelle et al., *Citoyen belge dans le système politique* (AGLOP) (Brussels: Institut Belge de science politique, 1975).

[17] There is one notable work that avoids all of the weaknesses of the modernization approach while highlighting many of its strengths. Young (fn. 13) addresses most of the issues not analyzed by the modernization theorists. He emphasizes the role of re-created identities in politicizing ethnicity in the Third World, mobilizing mass support, and defining the direction and scope of ethnic political activity. In addition, Young also borrows from cultural pluralist arguments in highlighting the social, economic, and political relations of subordination and domination among ethnic groups in defining the direction of ethnic conflict.

The weaknesses of the ethnic modernization approach and of modernization theory in general fed the heated intellectual debate that surrounded the internal colonialism approach of Michael Hechter; see Hechter, *Internal Colonialism: The Celtic Fringe in British National Development, 1536–1966* (Berkeley: University of California Press, 1975). Despite its novel approach to the relationship between modernization and ethnicity, Hechter's arguments suffer from the same weaknesses as the more mainstream modernization approaches.

many aspects of ethnic political activity were either poorly explained or left unexplained altogether. The ethnic politics literature of the 1970s had examined large economic and political changes and assumed that ethnic political activity would result from these changes. This approach tended to identify these changes as sufficient, rather than merely necessary, conditions for the rise of ethnic political movements. Just as the modernization approach had incorrectly assumed that Third World states would rapidly achieve economic parity with the early industrializers of the West, its ethnic derivative assumed the inexorable influence of modernization. Little attention was paid to the political obstacles faced by new ethnic movements. The conflictual modernization approach also did not fully examine how the processes of economic modernization and political development might result in different manifestations of ethnic political activity, depending on the balance of demographic, social, economic and political resources among ethnic groups within a state. It failed to draw any generalizable conclusions about what patterns of social, economic, and political change might result in certain manifestations of ethnic political activity.

Several scholars of the ethnic politics literature in the early 1980s expanded the conflictual modernization approach to address these deficiencies. It included not only the theoretical works discussed below, but also case studies such as the work of H. M. Drucker and Gordon Brown and that of Kenneth McRoberts and Dale Posgate.[18] This new literature owed much to cultural pluralist arguments, which highlighted relations of domination and subordination between ethnic groups under colonialism. Scholars such as John Sydenham Furnivall and Michael Garfield Smith examined states in which several different cultural groups tried to form a common society.[19] They argued that one cultural group will inevitably have to be politically and culturally dominant for a cross-cultural community to emerge. Their research focused on the importance of analyzing ethnic domination by examining the distribution of economic and political resources among ethnic groups.

In *Ethnopolitics: A Conceptual Framework*, Joseph Rothschild's conflictual modernization approach, the author introduces ethnic groups and the state as actors with economic and political resources at their disposal. Since the early 1960s ethnic political activity has been triggered by the differential impact of modernization on ethnic groups and by a decline

[18] Drucker and Brown, *The Politics of Nationalism and Devolution* (London: Longman, 1980); Kenneth McRoberts and Dale Posgate, *Quebec: Social Change and Political Crises*, rev. ed. (Toronto: McClelland Stewart, 1980).

[19] Furnivall, *Colonial Policy and Practice* (London: Cambridge University Press, 1948); Smith, *The Plural Society in the British West Indies* (Berkeley and Los Angeles: University of California Press, 1965).

in respect for the state. The latter has emboldened ethnic movements
challenge the state's legitimacy. However, says Rothschild, the form
existence of ethnic political activity is dependent on the balance amor
ethnic groups within a given state. Therefore, to argue that changes
the status quo, such as social mobility, will politicize ethnic identifica
tions, one must first understand the status quo, that is, the demographi
economic, and political balance among ethnic groups within a stat
Rothschild argues that these balances determine the resources and op
portunities available to ethnic groups in their political battles. The mos
common balance is a reticulate model where

> each ethnic group pursues a wide range of economic functions and occupations, and each economic class and sector organically incorporates members of several ethnic categories. But a certain amount of overrepresentation and underrepresentation of ethnic groups within economic classes and political power clusters is possible—indeed, likely. (pp. 81–82)

Within this reticulate system either the dominant or the subordinate group can politicize ethnic identifications, with strategy determined by the demographic, political, and economic balance as well as the choices of the ethnic leadership. Moreover, the state itself often uses its immense resources to politicize ethnic identifications.

Rothschild moves beyond earlier approaches by arguing that economic modernization and political development do not provide the necessary and sufficient conditions for the politicization of ethnic identifications. Instead the sufficient conditions for the rise of ethnic political movements are dependent on the economic, political, and ideological resources available to ethnic groups. Additional attention must be paid to the actual composition, traits, and cultural markers of an ethnic group, even if these groups were newly re-created by the forces of modernization.[20] In this, Rothschild borrowed from the resource mobilization school that studies political movements.[21] His approach holds out the hope of developing theories that emphasize the ethnic content of ethnic political activity. It could also move beyond a mere analysis of the general proclivity toward ethnic conflict and actually study the political obstacles to the achievement of the goals of ethnic political movements.

[20] Relations of subordination and domination among ethnic groups had been an important component of the cultural pluralist and Marxist writings on ethnicity, as well as of the writings of Crawford Young. In the early 1980s scholars sought to categorize these relationships in the hope that they would determine the resources and strategies of new ethnic political movements. The earlier theorists had used these relations mainly to determine the conditions for the existence of ethnic political identities.

[21] For a critical analysis of this school, see Thomas R. Rochon, "Political Movements and State Authority in Liberal Democracies," *World Politics* 42 (January 1990), 299–313.

Edward Tiryakian and Ronald Rogowski's *New Nationalisms of the Developed West: Toward Explanation* is another attempt to expand the modernization perspective and address these outstanding issues.[22] In "Causes and Varieties of Nationalism: A Rationalist Account," Rogowski uses rational choice theory to predict when ethnic identities will result in political conflict and to describe the ideological forms these conflicts will take. He emphasizes that modernization may create the necessary conditions for ethnic political activity but that the actual development of this conflict depends upon the balance of resources available to the various ethnic groups within a state. The modernization process determines the economic and political resources available to specific ethnic groups, and the relative levels of these resources in turn structure ethnic group ideologies, strategies, and political organization. Rogowski focuses both on "plural" states in which one ethnic group dominates another in a hierarchical cultural division of labor and on "pillarized" societies in which each ethnic group retains the skills necessary for the creation of a completely independent state. He argues that rational choice theory can explain how individuals from each type of ethnic group within a state react to ethnic group dominance—whether by assimilation, isolation, apathy, resistance, or minority nationalism. In addition, rational choice can also be used to explain the invention of a new ethnic identity or the option of nonethnic resistance to subordination.

Rogowski develops a series of hypotheses derived from his rational choice approach. In a state with a cultural division of labor and a stable system of social stratification, ethnic nationalism will be nonexistent. Upwardly mobile members of the subordinate ethnic group will lead a radical nationalist movement if they have not been assimilated by the dominant culture. This is the most prevalent form of nationalism, according to Rogowski; in the last fifty years it has appeared in Nazi Germany, India, Quebec, Scotland, and Africa (Rogowski and Tiryakian, 93). Conversely, upward mobility among the subordinates may result in a reactive nationalism among the nonelites of the dominant culture. This "redneck" nationalism has developed among Afrikaners, as well as among whites in Rhodesia and the American South and Protestants in Northern Ireland (p. 94). Downward mobility among members of the elite culture will not lead to nationalism unless these declining elites are expelled

[22] The volume by Tiryakian and Rogowski is divided into a theoretical/methodological part and comparative case studies. Many of the theoretical articles bear little relation to each other or focus only on certain specific aspects of ethnic movements. The case studies differ in their examination of religious, ethnic, and economic factors in Catalan, Scottish, Welsh, and Quebec nationalism. Rogowski's contribution is one of the few that lays out a parsimonious theory to account for the behavior of all ethnic movements in the developed West.

from the elite culture, at which point they lead the resistance of the su ordinate ethnic group.

Furthermore, a change in the mode of production in a state with cultural division of labor that "renders obsolete" the skills of the ethn elite may give rise to a nonelite, radical nationalism such as the natio alisms of eighteenth- and nineteenth-century Europe that followed t decline of landowning aristocracies (p. 95). If the changes in the mode production lead to the deskilling of the subordinate group, then the will not be a rise in nationalist activity because the subordinate grou will lack the skilled elites to direct the movement. If a change in the scal of production displaces an old elite and replaces it with a new one, the a radical nationalist movement will arise. However, if a cultural divisio of labor remains stable, then the pressure of a subordinate culture t assimilate would not, in and of itself, induce a subordinate nationalism In addition, changes in the terms of trade between a subordinate an dominant ethnic group will not result in nationalism unless thes changes improve the position of the nonelite and appear to be of a long duration. Then the resulting nationalism would be conservative, antire-volutionary, and nonviolent.

Following his description of nationalist possibilities in societies with a cultural division of labor, Rogowski turns to pillarized societies. If social mobility is much higher in one culture than in another in a pillarized society, either ethnic group may resort to manipulative nationalism, designed by traditional elites to maintain their position. This occurred in Quebec with the emergence of a conservative nationalism designed to preserve a rural parish culture as the province modernized between 1936 and 1960. If a new mode of production makes it costly to maintain traditional patterns of peaceful coexistence, ethnic conflict will ensue, with each ethnic group trying to assimilate its opponent's "inferior" culture. Rogowski believes this may have been the case in the conflict that developed between declining German Catholics and rising German Protestants with the growth of capitalism following the Reformation (p. 100). Under the status quo the elites of each ethnic group try to maintain social divisions between the two groups at the nonelite level so as to preserve a consociational society supported by two conservative nationalisms. However, if one culture decides to break that status quo either through assimilating the other culture or by destroying social unity, an elite-led defensive nationalism will develop within the threatened group.

Despite its contributions, Rogowski's typology suffers from some of the weaknesses previously emphasized in our review of the conflictual modernization approach. Essentially, his rational choice analysis turns upon the resources available to an ethnic elite: the existence of an ethnic

political movement depends on a sufficiently large ethnic elite with the requisite skills and resources to mobilize political support and sustain a movement. If this ethnic elite feels that a change in economic conditions opens a window of opportunity to expand its political power, it will act decisively, mobilizing an ethnic movement. Once again ethnic ideologies are resources to be mobilized at the behest of manipulative elites.

This conclusion is unsatisfactory from several perspectives. A rational choice theory should remain internally consistent. All participants should be acting in their own rational interest based on their own calculations. Even if supporters of an ethnic movement may eventually benefit from the movement, it appears that their decision to support the movement is rooted in an ideology formulated by an ethnic elite to further its own social and political positions. In fact, in some forms of ethnic nationalism this ideology is formulated against the interests of the mass of supporters. The dominant form of Francophone nationalism in Quebec between 1936 and 1960 encouraged a cultural nationalism that reinforced the dominance of Francophone political and religious elites while denying economic opportunities to the masses of Quebec Francophones. Thus, the acceptance of rational decision making for ethnic elites requires the denial of rational behavior on the part of their followers. From a methodological perspective this approach is therefore internally inconsistent. Essentially, Rogowski's argument suffers from the free-rider problem. Although it may appear rational for leaders to lead in order to obtain access to state economic and political resources, the followers need not follow. They do not directly benefit from ethnic nationalist agitation and moreover risk losing the economic resources they have. It is also doubtful that founders of radical nationalist movements can provide selective incentives to mobilize support before they have gained access to state resources.[23] From a theoretical perspective, Rogowski's reliance on the role of economic change in shaping an interest-laden ethnic political identity once again results in a description of ethnic nationalism that is manipulated by an elite.

A New Approach: The Psychological Dynamic

In response to the weaknesses of the dominant approaches to the study of ethnic politics, Donald Horowitz introduced a new approach in *Ethnic Groups in Conflict*. He holds a deep conviction that there is a unique

[23] Later in the volume Jack Brand argues that even electorally successful ethnic parties have trouble acquiring access to selective incentives. In his study of the rise of the SNP, Brand argues that the SNP could not have transformed governmental outputs into selective incentives for party members. All governmental programs were executed by the Scottish Office, an office over which the SNP had no control (Tiryakian and Rogowski, 282–83).

ethnic dynamic, some component of ethnicity that allows it to o whelm what are perceived to be rational economic and political inter In many Asian and African countries, for example, ethnic conflict see to dominate the political arena, even when the mobilization along eth lines is not in the interests of elites or followers. Horowitz emphasizes psychological dynamic that underlies the relationship between the caus of conflict, the development of an ethnic agenda, and the consequer permutations that these conflicts undergo. This emphasis on a psycho logical dynamic is an attempt to shift the focus of analysis from the struc tural conditions underlying ethnic conflict to a Weberian emphasis o *verstehen*: social, economic, and political changes result in ethnic conflic only as they are filtered through the mind of the political animal.

Horowitz seeks to develop a theory that will explain many aspects o ethnic conflict in Asia and Africa, including the severity of ethnic conflict; the forces that shape ethnic groups; the relationship between ethnic conflict and social class conflict; the goals and strategies of ethnic political movements; the likelihood of ethnic movements achieving their goals; the dynamics of ethnic party conflicts; the dynamics of ethnic military coups; and public policy alternatives for limiting ethnic conflict.

According to Horowitz, ethnic groups are constantly created and recreated anew. In establishing political boundaries in Asia and Africa, the colonial powers redefined the size and scope of ethnic groups. The expanded political boundaries led to the assimilation and differentiation of ethnic groups; as old groups disappeared, new ones emerged and others simply merged or split. The likelihood or intensity of ethnic conflict is therefore not dependent on particular aspects of ethnicity, such as race, language, and common ancestry. And no one trait is necessarily more incendiary than another in promoting ethnic tensions.

Horowitz is particularly attentive to the ability of ethnic conflict to overwhelm social class divisions; that is, he would argue, there is something unique about ethnic identity. Whereas individuals can overcome their social class and whereas social mobility is possible from generation to generation, ethnic ties are ascriptive. If one is born into an ethnic group, so be it; one cannot maneuver out of one's ethnic identity. Horowitz bases this argument on the notion that ethnicity is a comparative concept, and if the members of ethnic group B define someone as a member of ethnic group A, that individual is powerless to object. However, this argument flies in the face of the mutability argument, which presupposes that ethnicity is constantly being redefined. Why should B be able to define the ethnicity of A and A be unable to define his own ethnicity? Moreover, if A can redefine the ethnic identity of B, that should influence how B defines A.

According to Horowitz, the only way to study *ethnic conflict is by first determining* the social composition *of competing ethnic groups. In a ranked* ethnic system,

> mobility opportunities are restricted by group identity. In such systems, political, economic, and social status tend to be cumulative, so that members of [one] group are simultaneously subordinate in each of these ways to members of [another] group. Relations between the groups entail clearly understood conceptions of superordinate and subordinate status. (p. 22)

In such a system ethnic conflict will most likely take the form of a social revolution that emphasizes ideological divisions of class rather than ethnicity. However, in *unranked* systems conflict among ethnic groups is most likely to take the form of explicit ethnic conflict.

> In unranked systems, on the other hand, parallel ethnic groups coexist, each group internally stratified. Unlike ranked ethnic groups, which are ascriptively defined components of a single society, parallel groups are themselves incipient whole societies and indeed may formerly have constituted more or less autonomous whole societies. (p. 23)

Unranked ethnic systems have the potential for erupting into ethnic conflict because each ethnic group maintains a set of evaluations of other ethnic groups that defines stereotypical perceptions of other ethnic groups. These evaluations can be negative or positive in content and can reflect views of unranked ethnic groups as backward or advanced. Evaluations of backward groups include the terms lazy, traditional, submissive, and unintelligent. In turn the evaluations of advanced groups include enterprising, aggressive, arrogant, cunning, and miserly.

These evaluations had been advanced by the European colonizers to suit their own interests. Once verbalized and delineated, these perceptions of ethnic differences eventually found their way into governmental policy. Colonizers used different ethnic groups to divide and conquer, often selecting certain groups to staff their bureaucracies and armies. Their evaluations, based in group stereotypes, often became part of the nomenclature of the individual ethnic groups. Following the abandonment of the colony by the colonizer, certain political events highlighted these stereotypes and incited ethnic conflict. These conflicts erupted during the contest for preeminence in the independent state and were exacerbated by political events that required the delineation of boundaries and the dividing of resources among ethnic groups, such as a statewide census. At every stage of political development and national integration, then, each ethnic group invokes its own symbolic legitimacy in the polity.[24]

[24] In Parts II–V of the book Horowitz uses his typology of advanced and backward regions

Horowitz confronts many previously ignored issues. He argues t
there is a logic to ethnic conflict that influences the perceptions and a
tions of leaders and followers alike. This ethnic logic cannot be direct
translated into rational decision making, precisely because it is condi
tioned by a dynamic internal to the comparative nature of ethnic identit
and is not dependent on purely economic and political concerns. Thi
psychological dynamic feeds ethnic conflict at all stages and can determine the direction of ethnic conflict, as well as the strategies of ethnic political movements. Thus, once the causes of the rise of ethnic conflict have been explained, the agenda of political conflict is also described. By contrast, once the modernization theorists had described the causes of ethnic political conflict their job was complete.

There are certain limitations to Horowitz's hypotheses, however. Horowitz rejects the economic explanations for the rise of ethnic conflict in the Third World and, as a result, relies on the intervention of an external force—the colonizer—to disrupt the delicate balance between ethnic groups in the Third World. In unranked ethnic systems the perceptions developed and reinforced by the colonizer set in motion the dynamics of ethnic conflict. However, Horowitz realizes that he must explain how these perceptions maintain their credibility in the postcolonial era. After all, as Barrington Moore writes, "The assumption of inertia, that cultural and social continuity do not require explanation, obliterates the fact that both have to be recreated anew in each generation, often with great pain and suffering."[25] Consequently, Horowitz argues:

> The colonists set in motion a comparative process by which aptitudes and disabilities imputed to ethnic groups were to be evaluated. Those evaluations, solidly based in the group disparities that emerged, could not be dismissed as the irrelevant invention of a foreign overlord. Like the new polity and economy in which the disparities were embedded, the evaluations took hold. (p. 164)

There are both theoretical and practical problems with this argument. If these evaluations assume concrete expression in actual group disparities, why rely on the evaluations rather than on the disparities themselves to explain ethnic conflict? Horowitz's desire to disprove economic explanations for the rise of ethnic conflict has forced him to reject even the economic explanations that flow logically from his own theories. When

---

and advanced and backward ethnic groups in unranked systems to calculate the disposition of ethnic groups to secede, the possible patterns of ethnic party conflict, the militarization of ethnic conflict, and public policy alternatives for decreasing ethnic conflict.

[25] Barrington Moore, Jr., *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World* (Boston: Beacon Press, 1966), 486.

an advanced group describes a backward group as "poor," is the advanced group reinforcing a psychological prejudice or simply describing real economic differences that stand at the heart of an ethnic conflict? In addition, Horowitz argues that these evaluations are important in delineating ethnic conflict only in unranked systems, because in ranked systems ethnic conflict is subsumed under the rubric of social revolutions led by class conflict. However, if ethnic evaluations bring to life and reinforce the advanced and backward characteristics of various ethnic groups, why should they not result in ranked systems where the dynamic of conflict is class-based and not ethnically based? By focusing on the psychological dynamic of ethnicity, Horowitz echoes many of the earlier primordialist arguments while rejecting their logic. In addition, in a ranked system, such as exists in South Africa, can a convincing argument be made that revolutionary conflict is not also ethnic conflict? Finally, the introduction of a colonizer to upset the delicate balance among ethnic groups and hence re-create ethnic divisions helps little in the identification of the source of ethnic conflict in the developed world. What are the forces that can upset the balance between ethnic groups and re-create ethnic conflict in the industrialized world—where there is no overwhelming intervention by a colonizer that redefines the relationship among ethnic groups?

I believe that in decrying modernization theories regarding the rise and direction of ethnic political conflict, Horowitz has thrown out the baby with the bath water. Although the modernization approach has not yet yielded a satisfactory analysis of the internal dynamic of ethnic conflict, that does not mean that it does not have the potential to account for the development of ethnic political movements and the direction of ethnic conflict. If modernization is not a sufficient force to explain ethnic political conflict, it does nonetheless contribute to our understanding of it.

Re-creating "Primordial" Identities

The three most prevalent weaknesses of the modernization paradigm are all closely interrelated. By emphasizing the causes of the rise of ethnic movements, conflictual modernization theorists avoid detailed discussion of the ideological, strategic, and organizational apparatus of ethnic movements. And in emphasizing the economic and political interests of elites, theorists have avoided discussing the role of ethnic boundaries and ethnic cultural and ideological markers for determining the depth and breadth of movement support. However, these weaknesses are not an

of ethnic politics. One might argue that this difference in emphasis is rooted in the relative immaturity of the "religion in politics literature," which is still mired in the descriptive rather than the analytical. More accurately, it appears that the difference in emphasis is a function of the different organizational and symbolic character of religion and ethnicity.

Religious political movements that try to mobilize new political constituencies by appealing to a common religious tradition maintain certain tactical advantages over ethnic political movements. Organized religion provides a ready, integrated organizational structure and ecclesiastical leadership that does not flow from the existence of an ethnic group. In addition, religious political elites can often appeal to a common moral or ethical code. At different moments in history this moral code can be reinterpreted to attack social ills that require political rectification. Following the nineteenth-century waves of Irish and German immigration to the United States, for example, mainstream Protestants invoked their own moral codes, as defined by their religion, to mobilize support for prohibition. In the 1990s fundamentalists use the moral codes in their religious beliefs to mobilize in favor of the prohibition of abortion. Islam and Judaism provide rich codes of civil, religious, and criminal law that can be invoked to promote the integration of religious law into state legal systems. And indeed, in the Islamic world and in Israel, religious political movements have lobbied successfully to that end. A broad political agenda often flows from the religious beliefs upon which religious political movements are based.

The same cannot be said for ethnic political movements unless they also politicize religious symbols.[27] Although an ethnic identity is traditionally infused with a rich cultural, historical, and linguistic heritage, these symbols do little more than promote a psychological unity within an ethnic community. They neither create nor define a political agenda.[28] Thus, the willingness of students of religious politics to focus on religion is a reflection of the subject matter. The willingness of students of ethnic politics to avoid a discussion of ethnicity is a function of subject matter compounded by the constricting influence of the modernization perspective.

[27] At least in the industrialized world most ethnic political movements over the last twenty-five years have not politicized religious symbols. The forthcoming pages will elaborate on this point.

[28] The recent upsurge in demands for ethnoregional autonomy—the single focus of many ethnic movements—reinforces this observation. Once autonomy is considered by the national electorate, whether it is granted or not, the issue tends to lose its salience. This was the case with the autonomy movements in Scotland, Wales, Belgium, and Quebec in the 1970s and 1980s. Ethnicity did not provide a rich political agenda beyond the demand for regional autonomy.

Juan Linz has attempted to surmount these limitations in his article "From Primordialism to Nationalism" (Tiryakian and Rogowski, chap. 10). Linz examines the re-creation of ethnic identities and ethnic agendas in Spanish ethnoregions. He studies survey data from the Spanish and French Basque regions, Catalonia, Galicia, and Valencia to determine the forces that led to the structuring and re-creation of ethnic identities. Linz argues that traditional ethnic nationalist movements generally advocate primordial conceptions of nationalism rooted in common ancestry and the use of common ethnic instruments such as language. Over time, however, the processes of ethnic intermarriage and migration make these primordial sentiments increasingly anachronistic. As a result, ethnic nationalists turn to a territorial conception of ethnicity in their drive for regional autonomy that would be inclusive of those who do not fit the primordial qualifications of ethnicity. This conception of ethnicity inevitably creates a crisis of legitimacy within the ethnic movement. The movement is forced to shed its ethnic identity, which provides it with cultural roots and ties it to a historical past. Instead, the language of ethnic protest is limited to demands for regional autonomy. These demands have no unique ethnic character when compared with other autonomy movements. Consequently, the ethnic movement turns to primordial elements to reinforce its distinctiveness. Residents who are not members of the majority ethnic group within the region are forced to assimilate or challenge the ideology of the leadership of the ethnic movement. In the process, the creation of this tension between those who emphasize cultural identity and those who do not results in ethnic factionalism that over the long term may influence the internal ideological and organizational coherence of ethnic movements.

Unlike the earlier conflictual modernization theorists who only described the reasons for the emergence of ethnic conflict, Linz takes the conflictual modernization logic one step further. He is the first to argue that even if ethnic identities are re-created by the forces of social, economic, and political change, this process challenges ethnic movements in ways that might influence their strategies and tactics. This insight can tie the modernization approach to the ideological mobilization of ethnic political identities.

Linz has shown that an examination of the newly re-created ethnic identities can account for the role of mass support for ethnic political movements, describe the tensions embedded in ethnic ideologies, and help explain more than merely the forces that led to the initial rise of ethnic movements. The re-creation process can serve as an independent variable in describing the ideology and organization of ethnic move-

ments. This perspective can be highlighted by examining briefly the re-creation of ethnic identities in Quebec and Brussels. In both cases the process of re-creation led to the politicization of ethnic identifications and affected the direction taken by rising ethnic movements.

Expanding the Modernization Approach: The Re-creation Process

In the early 1960s many ethnic political movements rose to prominence in Western democracies. The most active political movements appeared in Canada, the United Kingdom, Belgium, Spain, Greenland, and Switzerland.[29] In several countries the leadership of these ethnic movements coalesced around political parties whose primary raison d'être was the improvement of the cultural, social, economic, and political conditions of their potential ethnic constituency.

The most successful of these ethnoregional parties appeared in Scotland, Quebec, and Belgium. The Scottish National Party in Scotland, the Parti Québecois in Quebec, the Volksunie in Flanders, the Rassemblement Wallon in Wallonia, and the Front Démocratique des Francophones in Brussels all arose as a result of electoral dealignments in their respective regions.[30] These dealignments were a result of the modernization process, namely, the declining importance of party social institutions, the communications revolution, the wide availability of educational opportunities, and social mobility. Dealignment provided the ethnoregional parties with large potential constituencies.

The ethnoregional parties were able to capitalize on dealignment because many industrialized democracies became increasingly concerned with regional political problems in the early 1960s. This new concern highlighted the ethnoregional parties' programs for regional autonomy.

The tensions embedded in the centralization and bureaucratization of the 1940s and 1950s set the conditions for the rise of regional concerns. Because of these tensions, the state and the traditional parties, in spite of their commitment to centralization, began to advocate regional policies.

It is not coincidental that in the 1940s and 1950s ethnoregional issues were not prominent in Western democracies. These states were centralizing political power and trying to mute the negative impact of unfet-

[29] Joseph R. Rudolph and Robert J. Thompson, "Ethnoterritorial Movements and the Policy Process: Accommodating Nationalist Demands in the Developed World" (Paper presented at the annual meeting of the Midwest Political Science Association, Milwaukee, Wisconsin, April 29–May 2, 1982), 1.

[30] See Newman (fn. 2), chap. 5.

tered economic forces. National governments created extensive social welfare schemes, regulated economic institutions, and stimulated economic growth, in the hope of limiting the debilitating consequences of the business cycle. The introduction of the mixed economy, the social welfare state, and full employment, as well as a commitment to economic growth, expanded the scope of centralized power. Regional governments merely helped execute national policies.

Nevertheless, it was precisely this centralization that provided the impetus for a new emphasis on regional concerns, for careful analysis by the central governments revealed the debilitating impact of regional economic problems. In the late 1950s and the early 1960s each of the dominant political parties offered its own proposals for solving these problems. Soon the state developed regional policies in response to the growing concern with regional problems. These policies intensified the attention that was paid to regional political issues.

Increased emphasis on regional issues served to re-create ethnic identifications in Scotland, Belgium, and Quebec, while dealignment created the vacuum in which these new identities could be politicized. This re-creation process not only led to the rise of new ethnic movements but also defined their ideologies and organizational structures. The concomitant processes of economic modernization and political centralization played key roles in determining the long-term success of these ethnic movements.

The combined influence of the modernization and centralization processes re-created ethnic identifications and ethnic organizations in Quebec and Brussels, thus determining the ultimate successes and failures of the Parti Québecois and the Front Démocratique des Francophones.[31] Prior to 1960 the dominant ethnic identity for Canadians of French origin was a French Canadian identity that manifested itself politically in the ideology of *la survivance*. The traditional rural parish culture of French Canada was glorified for its commitment to intellectual and cultural pursuits, whereas the materialistic, industrialized, secular Anglophone society was denigrated. The government of Quebec encouraged a French Canadian nationalism based on church domination of social and educational institutions and minimal governmental involvement in social and economic affairs. French Canadian nationalists saw Canadian confederation as a convenient institutional arrangement for maintaining their traditional parish social structure.

The political manifestation of *la survivance* became increasingly

[31] Of the five parties mentioned above, the PQ and FDF serve as the most striking examples of this re-creation, which affected all five parties.

anachronistic in the wake of postwar economic modernization in Quebec. In 1960 a new Liberal government in Quebec initiated a series of reforms designed to expand the scope of provincial social and economic programs. This era of reform, which lasted until 1966, came to be known as the Quiet Revolution. During and after that period the expansion in the size and powers of the Quebec provincial government had a definite effect on the ethnic identity of French Canadians. The Quebec government became the focus of Francophone aspirations for social and economic equality with Canadian Anglophones. French Canadian nationalism was transformed into Quebecois nationalism. The focus of ethnic identity shifted from the mythology of French Canadian parish culture to the Quebec state. Conservative nationalism that promoted political stagnation within confederation was swept away by demands for the expansion of the autonomous powers of Quebec. Both religious beliefs and ecclesiastical institutions were swept aside by a secular ideology centered on the role of the state in bringing Quebecois into the modern world. The idea that all French Canadians throughout Canada were connected by a common culture became passé. It was supplanted by a Quebecois identity that included only the Francophones of Quebec. It was these Francophones who could be "masters in their own house," possibly through independence for Quebec.[32] Thus, economic modernization broke down the old ethnic identities, while changes in state structures helped re-create them. This re-creation of ethnic identities in Quebec served as the ideological foundation for the growth of the PQ, an ethnic political party that advocated sovereignty for Quebec.

The process of economic modernization and political reforms also re-created ethnic identities in Brussels in the early 1960s. Belgium is divided into two ethnolinguistic regions, Dutch-speaking Flanders and French-speaking Wallonia. Ethnonationalist movements had traditionally erupted in Flanders and Wallonia, and the citizens of Brussels had traditionally identified with one ethnic region or the other on the basis of linguistic and ancestral ties.[33] However, a series of economic and legal changes in the early 1960s re-created a new Bruxellois identity for Francophones in Brussels.

[32] "Masters in our own house" was the slogan of the Parti Liberal du Québec during the key Quiet Revolution election of 1962.

[33] Brussels, the Belgian capital, is geographically located in Flanders and during the nineteenth century was a predominantly Dutch-speaking city. However, by the 1960s, 80% of the city's residents spoke French as their first language. Nevertheless, Brussels has always been legally recognized as a bilingual city. Since the 1950s the Francophone suburbs of Brussels have been extending into the countryside of Flanders, creating an "oil stain" that Dutch speakers have tried to limit through a linguistic "iron collar" around Francophone Brussels.

In response to the economic expansion of Francophone Brussels, which was perceived as a demographic threat by Dutch speakers, the government of Belgium passed a series of laws in the early 1960s to protect the bilingual status of the city. The boundaries of bilingual Brussels were drawn to exclude the Francophone suburbs of the city, which lay geographically within Flanders. Consequently, Francophones in certain areas of suburban Brussels were forced to educate their children in Dutch.

The impact of this legislation on politics in Brussels was profound. In response to the imposition of language options on the population, the city's Francophones responded by forming a new political party, the Front Démocratique des Francophones. The FDF initially advocated the protection of parental freedom of choice regarding the language of a child's education. Within four years the FDF began to argue that the best way to defend this freedom of choice was to achieve regional autonomy for Brussels. The new FDF constituency was an amalgam of citizens who traced their ancestry to residents of Brussels, as well as to Flemings and Walloons. Some supporters of the FDF had been born in Flanders and moved to Brussels. Their new common identity was that of French speakers in Brussels. They were Bruxellois but certainly not Walloons. Thus, the linguistic legislation of the early 1960s paved the way for the creation of a new Bruxellois identity and the rise of a powerful ethnic political movement.

For the students of ethnic politics from the modernization school this re-creation of ethnic identities does not come as a surprise. Although they have traditionally emphasized the role of economic modernization rather than political centralization and regionalization in the re-creation process, they would agree that this process paved the way for the ethnic resurgence. But the characteristics of this new ethnic identity have not been perceived as directly relevant to the evolution of ethnic political conflict. This is where the conflictual modernization paradigm has been undersold. The re-created identity had become an important independent variable in determining the evolution of conflict.

Scholars have often asked how politicized ethnic identifications can overcome the traditional social and economic divisions in most polities. Clearly, the re-creation of ethnic identities has opened the door to ethnic political movements by raising issues not addressed by the traditional political elites. In Quebec and Brussels, in particular, the PQ and FDF both emphasized issues of regional autonomy that were not addressed by the major political parties. As a result, these parties cut across the old

cleavage lines and formed new electoral coalitions. However, the re-creation of ethnic identities also built political limits into the success of these movements.

The new ethnic political movements sprang from the ashes of the devastated old ethnic identities. The Quebecois nationalist movement was the antithesis of French Canadian nationalism. The Bruxellois movement presupposed the denial of Flemish and Walloon identities. This has more than just rhetorical significance. The elites and institutions that had fed earlier ethnic identities were now swept aside and made irrelevant. The resources and institutions of French Canadian nationalism had been decimated and were of no political use to the PQ. Similarly, the FDF could not tap any social or political institutional base. In contrast to religious movements, which might have vast institutional networks to tap, the new ethnic movements have to form an organizational base from the bottom up and therefore lack the patronage network that gives supporters strong social links to the parties. It is quite difficult to overcome the long-established social and economic networks of the traditional political parties without an apparatus of one's own.

Instead of relying on existing institutional networks, the new ethnic political movements have to rely on ideological legitimacy to cement their constituency. However, as Linz has noted, these new ideological legitimacies lack primordial roots or anything else that underscores the cultural uniqueness of the ethnic group. Instead they are territorial ethnicities that promote regional autonomy or independence. The ethnic agenda of autonomy is a limited one that can be realized with a clearly defined legislative package, but once it has been acted upon it leaves the ethnic movement without a raison d'être. Indeed, this is precisely what has happened in most cases involving the rise of ethnic movements in industrialized democracies. For example, the referendum defeats on sovereignty-association for Quebec and on devolution for Scotland led to the decline of ethnic movements in these regions. In essence the very re-creation of ethnic identities has created the potential for ethnic movements but has left them without competitive institutional and ideological resources.

This analysis allows us to reevaluate the validity of the conflictual modernization approach. Clearly, social, economic, and political modernization can account for more than just the necessary conditions for the rise of ethnic conflict. The manner in which modernization re-creates and politicizes ethnic identifications may also determine the ideological and organizational resources available to ethnic movements. It is the distribution of these resources that determines the success of ethnic

movements and the future direction of ethnic conflict. In industrialized democracies modernization often creates the conditions for ethnic conflict while destroying the links between an ethnic movement and its ideological and organizational heritage. In authoritarian states, where the distribution of these resources is determined more by state power than by the process of economic modernization, ethnic conflict rises and falls as a function of the government's ability to redistribute political resources.

Several tests of this approach could be found in countries in which ethnic movements have arisen following moves toward democratization. Such case studies might include the ethnic autonomy movements in Spain and the Baltic Republics. The rise of these ethnic movements cannot easily be attributed to conflictual modernization. Economic modernization in the decades prior to the death of Franco and the emergence of *perestroika* was minimal. Rather, the collapse of mobilizing political parties in these ethnic regions led to a political dealignment that provided facilitating conditions for the emergence of new social movements. Furthermore, the high degree of centralization of the authoritarian and totalitarian states and its differential impact on various ethnic regions fueled ethnic resentments.

Although the redistribution of resources by the state created the conditions for the rise of these ethnic movements, it also deprived them of their traditional romantic cultural symbols and ethnic organizations. As a result, the Basque, Catalan, Lithuanian, Estonian, and Latvian movements are in a strategic paradox. They are at once re-creating an older romantic nationalist identity and forging a newer political identity. The romantic nationalist identity has been stripped of much of its organizational structure by years of centralizing dictatorship. Meanwhile, emphasis on the economic, as opposed to cultural, benefits of autonomy creates a set of expectations that ethnic leaders will have as much success of fullfilling as Mikhail Gorbachev has had. The success of the Baltic movements is dependent upon overcoming this strategic problem, a problem the Catalans and Basques have been wrestling with for well over a decade.

Conclusion

An argument that emphasizes the ethnic component of ethnic movements, rather that merely the large economic and political changes that precede the creation of ethnic movements, provides meaningful insights into the study of the direction of ethnic political conflict in both the in-

dustrialized and the developing worlds.[34] By emphasizing the ideological component of ethnicity, it addresses the reasons for mass participation in ethnic movements, brings an ethnic component to the study of ethnic politics, and expands the purview of the conflictual modernization approach. We are not advocating the abandonment of the modernization paradigm in favor of a social movement approach that emphasizes resources, ideologies, and mobilization tactics over macroeconomic and political forces. Rather, we argue that the two should be combined. Only through a proper interpretation of the conflictual modernization approach—one that emphasizes the relationship between macroeconomic and political changes and their influence on specific political movements and institutions—can we live up to the promise of the modernization paradigm, create an integrated theory of ethnic conflict, and address many of the essential questions that most students of ethnic conflict have not yet considered. The resurgence of ethnic conflict in the communist world now makes this task imperative.

[34] In the Third World, the paucity of opposing institutional networks and integrated ideological systems has meant that even newly re-created ethnic movements might be able to survive the obstacles that bring down ethnic political movements in the developed world.

PUBLIC OPINION, DOMESTIC STRUCTURE, AND FOREIGN POLICY IN LIBERAL DEMOCRACIES

By THOMAS RISSE-KAPPEN*

INTRODUCTION

HOW is it to be explained that similarly powerful states more often than not respond differently to the same international conditions and constraints? Efforts to answer this question lead inevitably to the study of the domestic sources of foreign policy and international politics. One issue, though, is rarely addressed by the available literature: Who is in charge of the foreign policy-making process in liberal democratic states? Elites or masses? Who influences whom? What is the policy impact of societal actors and public opinion? Are public attitudes on foreign affairs manipulated by the elites? Finally, if public attitudes follow similar patterns, what accounts for differences between various countries in the policy impact of public opinion?

This paper tries to shed light on the relationship between public opinion and the foreign policy of liberal democracies.[1] I will argue that the policy impact of public opinion does not depend so much on the specific

* Drafts of this paper were presented at the European Consortium of Political Research (ECPR) Workshop on "Public Opinion, Foreign Policy, and the Democratic Process," Bochum, West Germany, April 2-7, 1990, and at the Annual Convention of the International Studies Association, Washington D.C., April 10-14, 1990. For very valuable comments I thank Richard Eichenberg, Matthew Evangelista, Philip Everts, Kjell Goldmann, Paul Joseph, Claudia Kappen, Peter Katzenstein, Richard N. Lebow, David Meyer, Nobuo Okawara, Hans Rattinger, Bruce Russett, and the participants in the ECPR workshop. For their help and advice regarding public opinion data I thank Lisa Brandes and Richard Eichenberg.

[1] The paper is part of a growing body of literature. See, for example, Michael Barnett, "High Politics Is Low Politics: The Domestic and Systemic Sources of Israeli Security Policy, 1967-1977," *World Politics* 42 (July 1990), 529-62; Richard Eichenberg, *Public Opinion and National Security in Western Europe* (Ithaca, N.Y.: Cornell University Press, 1989); Matthew Evangelista, *Innovation and the Arms Race* (Ithaca, N.Y.: Cornell University Press, 1988); Gert Krell, *Rüstungsdynamik und Rüstungskontrolle: Die gesellschaftlichen Auseinandersetzungen um SALT in den USA 1969-1975* (Frankfurt am Main: Haag und Herchen, 1976); Richard N. Lebow, *Between Peace and War* (Baltimore, Md.: Johns Hopkins University Press, 1981); Jack S. Levy, "Domestic Politics and War," *Journal of Interdisciplinary History* 18 (Spring 1988), 653-73; Erwin Müller, *Rüstungspolitik und Rüstungskontrolle: Fall USA* (Baden-Baden: Nomos, 1985); Bruce Russett, *Controlling the Sword: The Democratic Governance of National Security* (Cambridge: Harvard University Press, 1990); and Eugene Wittkopf, *Faces of Internationalism: Public Opinion and American Foreign Policy* (Durham, N.C.: Duke University Press, 1990).

issues involved or on the particular pattern of public attitudes as on the *domestic structure* and the *coalition-building processes* in the respective country. The paper analyzes the impact of public opinion on the foreign policy-making process in four liberal democracies with distinct domestic structures: the United States, France, the Federal Republic of Germany and Japan. I will show that differences in political institutions, policy networks, and societal structures account for different foreign policy outcomes when the impact of the international environment is controlled for and public attitudes follow similar patterns across countries. Thus the four countries responded differently to Soviet policies during the 1980s despite more or less comparable trends in mass public opinion. The analysis of the interaction between public opinion and elite coalition-building processes in the four countries reveals that the policy outcomes differ according to variances in domestic structures and not in the international status of the states.

Public Opinion and Foreign Policy: Theoretical Concepts and Methodological Problems

Who Follows Whom? Elites and Masses

Most of the available literature on the interaction between mass public opinion and elites in the foreign policy-making process of liberal democracies can be categorized according to two broad concepts.[2] In accordance with the pluralist theory of democracy, a "bottom-up" approach assumes that the general public has a measurable and distinct impact on the foreign policy-making process.[3] In sum, leaders follow masses. However, the bottom-up concept is difficult to reconcile with the following empirical counterevidence:

—In Western Europe, the U.S., and Japan, elites and masses show similar support for basic foreign policy goals and institutions. Prevailing political, religious, and ideological cleavages structure elite as well as mass public opinion.[4]

[2] For excellent reviews of the debate, see Philip Everts and Arthur Faber, "Public Opinion, Foreign Policy, and Democracy" (Paper presented at the ECPR Workshop "Public Opinion, Foreign Policy, and the Democratic Process," Bochum, April 2–7, 1990); and Russett (fn. 1).

[3] See, e.g., Benjamin I. Page and Robert Y. Shapiro, "Effects of Public Opinion on Policy," *American Political Science Review* 77, no. 1 (1983), 175–90.

[4] For Western Europe, see Eichenberg (fn. 1), pt. 5. For the U.S., see Ole R. Holsti and James N. Rosenau, "The Domestic and Foreign Policy Beliefs of American Leaders," *Journal of Conflict Resolution* 32 (June 1988), 248–94; and Wittkopf (fn. 1). For Japan, see Davis Bobrow, "Japan in the World: Opinion from Defeat to Success," *Journal of Conflict Resolution* 33 (December 1989), 571–603.

—There are many cases in which crucial foreign policy decisions have been taken in the absence of mass public consensus. Examples are the U.S. decisions in favor of an active international role in the postwar world and of becoming permanently involved in European security affairs; the West German decisions to rearm and join NATO in the early 1950s and to pursue an active *Ostpolitik* in the late 1960s; the French decisions to build an independent nuclear force in the 1950s and to leave NATO's military institutions in the mid-1960s.

Therefore, a second approach representing the conventional wisdom in the literature suggests a "top-down" process, according to which popular consensus is a function of the elite consensus and elite cleavages trickle down to mass public opinion. This viewpoint concurs with either a "power elite" (C. Wright Mills) or a state-centered and realist approach to foreign policy. It is assumed that the public is easily manipulated by political leaders, because of (1) the low salience, or significance, of foreign and security policy issues as compared with economic policies, (2) the low degree of knowledge about the issues involved, and (3) the volatility of public opinion.[5]

Again, empirical evidence suggests that these assumptions are questionable:

—While only a minority can be regarded as politically active, large portions of the public seem regularly to follow news about foreign policy in the media.[6] While domestic problems usually outweigh foreign and security issues in public salience, data reveal that substantial minorities consider foreign affairs among the most important problems facing their respective countries. On the average, 20–30 percent of the public indicate serious concern about foreign affairs.[7]

[5] See, e.g., Noam Chomsky and Edward Herman, *Manufacturing Consent* (New York: Pantheon, 1988); Benjamin Ginsberg, *The Captive Public: How Mass Opinion Promotes State Power* (New York: Basic Books, 1986); Barry Hughes, *The Domestic Context of American Foreign Policy* (San Francisco: Freeman, 1979); Michael Margolis and Gary Mauser, eds., *Manipulating Public Opinion: Essays on Public Opinion as a Dependent Variable* (Belmont, Calif.: Wadsworth, 1989); and James N. Rosenau, *Public Opinion and Foreign Policy* (New York: Random House, 1961).

[6] During election campaigns in the U.S., for example, the majority of the public was aware of the different foreign policy choices under debate and could identify the positions of the presidential candidates. See John H. Aldrich et al., "Foreign Affairs and Issue Voting: Do Presidential Candidates 'Waltz before a Blind Audience'?" *American Political Science Review* 83, no. 1 (1989), 123–41; Hans Rattinger, "Foreign Policy and Security Issues as Determinants of Voting Behavior in the 1988 U.S. Presidential Election" (Paper presented at the ECPR Workshop "Public Opinion, Foreign Policy, and the Democratic Process," Bochum, April 2–7, 1990). See also Thomas Graham, "The Pattern and Importance of Public Knowledge in the Nuclear Age," *Journal of Conflict Resolution* 32, no. 2 (1988), 319–34.

[7] For the U.S., see Tom W. Smith, "The Polls: America's Most Important Problem, Part I: National and International," *Public Opinion Quarterly* 49, no. 2 (1985), 264–74; for France, West Germany, and Japan, see Hastings and Hastings, eds., *Index to International Public Opinion, 1982–1983* (New York: Greenwood, 1984), 563; *and 1983–1984*, 689.

—Public attitudes, at least on basic foreign policy issues, seem to be more stable than is usually assumed. A large percentage of the public does not change attitudes frequently. In sum, the public seems to be more rational and less open to elite manipulation than the top-down thesis asserts.[8]

—That leaders try to manipulate masses and that governments launch sophisticated propaganda campaigns only suggest that the elites take the power of the uneducated masses seriously and feel vulnerable to it. The very existence of state propaganda and efforts at "spin control" belies the "power elite" hypothesis.

In addition to these empirical problems, the bottom-up and the top-down models both suffer from conceptual shortcomings. First, they treat masses and elites as unitary actors. While certain segments of the public may be manipulated by government propaganda, others may resist efforts to influence them.[9] One should at least distinguish between (1) *mass public opinion*, (2) the *attentive public*, which has a general interest in politics, and (3) *issue publics*, which are particularly attentive to specific questions. Moreover, elites are themselves frequently divided, and different segments of the elite try to convince the general public of their respective viewpoints.[10]

Second, a simplistic view of either bottom-up or top-down processes tends to ignore that public opinion and societal groups may influence the policy-making process in several ways and at different stages. They can directly affect the choices of top decision makers by changing policy goals or how those goals are prioritized, by narrowing the range of options and/or means to implement goals, or by winning symbolic concessions in the sense of changed rhetoric rather than policy reforms. Moreover, the public may also indirectly affect policies by influencing the coalition-building processes among the elites. It can strengthen or weaken the positions of bureaucracies or single actors within the government. Public

[8] Robert Shapiro and Benjamin I. Page, "Foreign Policy and the Rational Public," *Journal of Conflict Resolution* 32, no. 2 (1988), 211–47. For data, see Eichenberg (fn. 1); Gregory Flynn and Hans Rattinger, eds., *The Public and Atlantic Defense* (London: Croom Helm, 1985); Thomas Graham, *American Public Opinion on NATO, Extended Deterrence, and Use of Nuclear Weapons* (Cambridge: Center for Science and International Affairs, Harvard University, 1989); and Russett (fn. 1), chap. 4.

[9] For example, the well-documented "rally around the flag" effect, according to which presidents can use military interventions to raise their popular support, applies only to a minority of the U.S. population and does not last very long. See Russett (fn. 1), 34–51. For the original study on the effect, see John Mueller, *War, Presidents, and Public Opinion* (New York: Wiley, 1973). For a general argument regarding elite-public interactions, see Paul Joseph, *Realizing the Peace Dividend: The End of the Cold War and the Emergence of Progressive Politics* (forthcoming), chap. 4.

[10] For a discussion of the cleavages in U.S. public and elite opinion, see Michael Mandelbaum and William Schneider, "The New Internationalism: Public Opinion and American Foreign Policy," in Kenneth Oye et al., eds., *Eagle Entangled: U.S. Foreign Policy in a Complex World* (New York: Longman, 1979), 34–90; and Wittkopf (fn. 1).

nion and the activities of public interest groups may lead to changes /or realignments within or between political organizations such as ties. These indirect effects on policy are difficult to trace and are refore easily overlooked. However, their impact may be as important he above-mentioned direct influences.

'inally, it cannot be assumed that public and elite opinion interact h each other and are transformed into policy decisions in the same r across different countries. In other words, similarities in public at-des across various countries do not necessarily lead to similar policies. lic attitudes and policies toward nuclear weapons provide a good mple. The patterns of mass public opinion in the U.S., Western Eu-e, and Japan are more or less similar. Nuclear weapons are mostly rated for their deterrent value, provided that arms control efforts are lerway. Attitudes become increasingly hostile if nuclear use in war-e situations is contemplated. Throughout the cold war, "first use" of lear weapons had only minority support in the U.S., France, West many, and Japan.[11]

Iowever, for all these similarities in public attitudes, the nuclear pol-s of the four countries have differed from one another. Japan has had strongest antinuclear policy in place since the late 1960s. The Federal ublic of Germany changed from an outspoken supporter of U.S. ex-led deterrence and "coupling" to a strong advocate of nuclear arms trol. The United States emphasized arms control in the 1970s, turned massive arms buildup during Ronald Reagan's first term, and re-ed to arms control from the mid-1980s on. Finally, France has pur-l the most pronuclear policy of the four countries and has refused so to take part in nuclear arms control.

'he variations in policies are in part a result of different positions and abilities in the international environment (for example, the nuclear us of France and the U.S. as compared with that of Japan and Ger-ny). However, while the frequent changes in United States arms-rol policies might be considered to be reactions to changes in the 'eat environment" (Soviet arms buildup in the 1970s; Soviet interven- in Afghanistan; the "Gorbachev revolution"), French policy seems emain largely unaffected by either external or internal factors. The ige in West German policies even runs counter to what one would ect from looking solely at the "threat environment" such as the Soviet

'For data regarding the U.S., see Graham (fn. 8); and Russett (fn. 1), 57–65; regarding ern Europe, see Eichenberg (fn. 1), chap. 4; and Stephen Szabo, "West European Public 'ptions of Security Issues" (Manuscript, prepared for the USIA, July 1988); for Japan, obrow (fn. 4).

arms buildup of the late 1970s. The stability of Japanese nuclear policie: while in line with public attitudes, is also not explicable in terms o changes in the country's international environment.

The Missing Link: Domestic Structures and Coalition-building Processes

If, first, public attitudes as such cannot account for differences in policie and, second, variations in the international environment do not explair them either, what is the missing link between mass public opinion anc elite decisions in foreign and security policy? It is suggested here that on has to look at *domestic structures* and *coalition-building processes* to under stand the impact of public opinion on the foreign policy of liberal de mocracies.[12]

Domestic structure approaches deal with the nature of the politica institutions (the "state"), basic features of the society, and the institu tional and organizational arrangements linking state and society anc channeling societal demands into the political system. In other words domestic structures determine how political systems respond to societa demands. Social movement research talks about "political opportunity structures."[13]

Several approaches can be distinguished. The first focuses on *state institutions* and has found its most prominent expression in the concept of "strong" and "weak" states. It emphasizes the degree of centralization of state institutions and the ability of political systems to control society and to overcome domestic resistance.[14] Weak states have fragmented political institutions and are open to pressures by societal interest groups and political parties. Their ability to impose policies on society and to extract resources from it is fairly limited. Strong states, by contrast, con-

[12] For a similar argument focusing on the effects of party systems and electoral rules, see Eichenberg (fn. 1), 235–41.

[13] See, for example, Herbert P. Kitschelt, "Political Opportunity Structures and Political Protest: Anti-Nuclear Movements in Four Democracies," *British Journal of Political Science* 16 (January 1986), 57–85; and Sidney Tarrow, *Struggle, Politics, and Reform: Collective Action, Social Movements, and Cycles of Protest* (Ithaca, N.Y.: Center for International Studies, Cornell University, 1989). For an overview, see David Meyer, "Peace Movements and National Security Policy: An Agenda for Study" (Paper prepared for the Annual Meeting of the American Political Science Association, Atlanta, Ga., August 31–September 3, 1989).

[14] The concept was originally developed by Peter Katzenstein, "International Relations and Domestic Structures: Foreign Economic Policies of Advanced Industrial States," *International Organization* 30 (Winter 1976), 1–45; Stephen Krasner, *Defending the National Interest: Raw Materials, Investments, and U.S. Foreign Policy* (Princeton: Princeton University Press, 1978). For more recent discussions, see Michael Mastanduno et al., "Toward a Realist Theory of State Action," *International Studies Quarterly* 33 (December 1989), 457–74; and G. John Ikenberry, "Conclusion: An Institutional Approach to American Foreign Economic Policy," *International Organization* 42 (Winter 1988), 219–43.

sist of centralized political institutions with strong bureaucracies; they are able to resist public demands and to preserve a high degree of autonomy vis-à-vis society.

However, the "strong versus weak states" distinction, while parsimonious, is too simplistic to account for the variations between domestic structures. Weak states like the U.S. are sometimes able to conduct highly efficient policies, whereas strong systems might not always pursue forceful and energetic foreign policies.[15] Moreover, institutionalist approaches have been challenged as apolitical and therefore unable to explain specific policy outcomes. Rather than emphasizing state structures, one should analyze the *coalition-building processes* within societies and political systems.[16] These approaches focus on the "policy networks," that is, the mechanisms and processes of interest representation by political parties and interest groups that link the societal environment to the political systems. This concept emphasizes the ability of political actors to build consensus among the relevant elite groups in support of their policies.[17]

Simply to set off the coalition argument against the institutionalist approach seems to be inappropriate, though. On the one hand, state structures do not determine the specific content or direction of policies. On the other hand, coalition building takes place in the framework of political and societal institutions. The structures of the society and the political system determine the size and strength of policy coalitions needed to create the support basis for specific policies. Domestic structures also seem to account for general features of foreign policies, the degree of stability as well as the level of activity and commitment.[18]

In sum, a "mixed" approach encompassing both institutional structures and coalition-building processes seems to be more appropriate. In

[15] See, for example, G. John Ikenberry, "The Irony of State Strength: Comparative Responses to the Oil Shocks in the 1970s," *International Organization* 40 (Winter 1986), 105–37. See also the special edition of *International Organization* 42 (Winter 1988), "Approaches to Explaining American Foreign Economic Policy."

[16] Peter Gourevitch, "The Second Image Reversed: The International Sources of Domestic Politics," *International Organization* 32 (Autumn 1978), 881–911, at 904.

[17] For coalition-building approaches, see, e.g., Ernst-Otto Czempiel, *Amerikanische Außenpolitik* (Stuttgart: Kohlhammer, 1979); Joe D. Hagan, "Regimes, Political Oppositions, and the Comparative Analysis of Foreign Policy," in C. F. Hermann et al., eds., *New Directions in the Study of Foreign Policy* (Boston: Allen and Unwin, 1987), 339–65; Helen Milner, *Resisting Protectionism: Global Industries and the Politics of International Trade* (Princeton: Princeton University Press, 1988); Thomas Risse-Kappen, *Die Krise der Sicherheitspolitik: Neuorientierungen und Entscheidungsprozesse im politischen System der Bundesrepublik Deutschland, 1977–1984* (Mainz-Munich: Grünewald-Kaiser, 1988); and Ronald Rogowski, *Commerce and Coalitions: How Trade Affects Domestic Political Alignments* (Princeton: Princeton University Press, 1989).

[18] See Hagan (fn. 17).

their later works Peter Gourevitch and Peter Katzenstein, who had previously argued about the respective merits of institutionalist versus coalition-building concepts, both moved toward combined approaches that basically emphasized three factors:[19]

1. The nature of the *political institutions* and the degree of their centralization: Is executive power concentrated in the hands of one decision maker (president, prime minister, chancellor) who controls the bureaucratic infighting among governmental agencies? To what extent can the government control the legislative process?

2. The *structure of society* regarding its polarization, the strength of social organization, and the degree to which societal pressure can be mobilized: How heterogeneous is the society in terms of ideological and/or class cleavages? How well developed are social coalitions and organizations in their ability to express grievances and raise demands?

3. Finally, the nature of the coalition-building processes in the *policy networks* linking state and society:
   a. In countries with centralized political institutions but polarized societies and rather weak social organizations, the policy network is likely to be *state-dominated*. The policy-relevant coalition building would then be restricted to the political elites and would more or less exclude societal actors and/or public opinion.
   b. By contrast, *societal control* of the policy network is to be expected in countries with comparatively homogeneous societies and a high degree of societal mobilization but weak state structures. The policy-relevant coalition building would take place among societal actors; accordingly, public opinion would play a major role.
   c. Countries with political institutions and social organizations of comparable strength are likely to have a policy network characterized by *democratic corporatism*. Political and societal actors would be engaged in continuous bargaining processes in search of policy compromises in an environment of give-and-take.[20] As a result, some sorts of middle-of-the-road-policies are to be expected, reflecting the common denominator of public opinion.

The Impact of Public Opinion in Four Liberal Democracies

The United States, France, Germany, and Japan are characterized by distinct domestic structures. I argue that these differences largely account

[19] See Peter Gourevitch, *Politics in Hard Times* (Ithaca, N.Y.: Cornell University Press, 1986); Peter Katzenstein, "Introduction" and "Conclusion," in Katzenstein, ed., *Between Power and Plenty: Foreign Economic Policies of Advanced Industrial States* (Madison: University of Wisconsin Press, 1978), 3–22, 295–336; and idem, *Small States in World Markets* (Ithaca, N.Y.: Cornell University Press, 1985).

[20] For the concept of "democratic corporatism," see Peter Katzenstein, *Corporatism and Change* (Ithaca, N.Y.: Cornell University Press, 1984); and Katzenstein (fn. 19, 1985). See also Philippe Schmitter and Gerhard Lehmbruch, eds., *Trends towards Corporatist Intermediation* (Beverly Hills, Calif.: Sage, 1979).

for the variations in the impact of public opinion on foreign and security policy. And I test this proposition by discussing the various responses to changes in Soviet foreign policy from the late 1970s to the late 1980s (from Brezhnev to Gorbachev).

The Domestic Structures of the United States, France, Japan, and the Federal Republic of Germany

POLITICAL INSTITUTIONS

The U.S. undoubtedly has the most decentralized foreign and security policy-making structure of the four countries. The built-in tensions within the executive between the Pentagon, the State Department, and the National Security Council, which lead to continuous infighting over defense and arms-control policies, are well documented. Moreover, Congress has more authority over the conduct of foreign policy than do most other Western parliaments due to (1) the weakness of the American party system, which severely limits the power of the executive over congressional decisions and (2) institutional provisions such as the two-thirds majority requirement for the ratification of international treaties.

The very different French political structure consists of the strongest state institutions of the four countries. The Fifth Republic institutionalized a centralized political system that all presidents since de Gaulle have reinforced. The power of the French bureaucracy enhances the strength of the executive.[21] This general feature of the French political system is particularly relevant for foreign and defense policy-making, the *domaine réservée* of the president.[22] Furthermore, the French Parliament—in striking contrast to the U.S. Congress—plays an almost negligible role in foreign policy. In sum, the centralization of the French decision-making apparatus seems to be even greater in foreign policy than in other issue-areas.

Japan resembles the French case insofar as its political system is usually described as a strong state. Indeed, the power of the state bureaucracy extends into foreign affairs. However, the Foreign Ministry, which dominates Japanese foreign and security policy and outweighs the De-

[21] See Vincent Wright, *The Government and Politics of France* (New York: Holmes and Meier, 1978). On the role of the bureaucracy in French foreign policy, see Mark B. Hayne, "The *Quai d'Orsay* and the Formation of French Foreign Policy in Historical Context," in Robert Aldrich and John Connell, eds., *France in World Politics* (London: Routledge, 1989), 194–218.

[22] See Alfred Grosser, *Affaires extérieures: La politique de la France, 1944–1984* (Paris: Flammarion, 1984), 151–54. For a critical analysis of the institutions of French foreign policy, see Samy Cohen, *La monarchie nucléaire: Les coulisses de la politique étrangère sous la Ve République* (Paris: Hachette, 1986).

fense Agency, usually cannot compete internally with the Ministry of International Trade and Industry (MITI) or the Ministry of Finance.[23] The distribution of power within the government thus reflects the predominance of economic over defense issues in Japanese foreign policy. While the Japanese parliament, the Diet, plays only a minor role in foreign affairs, the Liberal Democratic Party (LDP), which has controlled the government since 1955, is among the most important actors in Japanese foreign policy.[24]

In the Federal Republic of Germany executive control over foreign and defense policy is generally stronger than in the U.S.[25] With the exception of the 1972 ratification process concerning the treaties with Eastern Europe and the Soviet Union (*Ostpolitik*), the role of Parliament has been fairly limited. This weakness, however, does not simply result in strengthening executive power as in France. Rather, the political parties constrain both the legislative *and* the executive branches. Partly as a result of peculiar election procedures,[26] the party system is smaller and less polarized than in France. The ruling parties in coalition governments must engage in constant consensus building on major foreign policy decisions. "Party democracy" permeates the government bureaucracy insofar as internal divisions since the mid-1960s are usually related to the party affiliation of the respective minister rather than to traditional bureaucratic roles.[27]

STRUCTURE OF THE SOCIETY

In sharp contrast to its centralized political system, French society is usually described as fragmented along ideological, religious, and class lines.[28]

[23] See Reinhold Drifte, *Japan's Foreign Policy* (London: Chatham House Papers, Royal Institute of International Affairs, 1990), 21–24; Haruhiro Fukui, "Policy-making in the Japanese Foreign Ministry," in R. A. Scalapino, ed., *The Foreign Policy of Modern Japan* (Berkeley: University of California Press, 1977), 3–35; and Gaston J. Sigur, "Power, Politics, and Defense," in J. H. Buck, ed., *The Modern Japanese Military System* (Beverly Hills, Calif.: Sage, 1975), 181–95.

[24] For details, see Hans Baerwald, "The Diet and Foreign Policy," in Scalapino (fn. 23), 37–54; and John Welfield, *An Empire in Eclipse: Japan in the Postwar American Alliance System* (London: Athlone Press, 1988).

[25] See Helga Haftendorn, ed., *Verwaltete Außenpolitik* (Cologne: Wissenschaft und Politik, 1978); and Barry Blechman et al., *The Silent Partner: West Germany and Arms Control* (Cambridge, Mass.: Ballinger, 1988), 27–60. On the German domestic structure in general, see Peter Katzenstein, *Policy and Politics in West Germany* (Philadelphia: Temple University Press, 1989).

[26] In particular, the proportionate vote combined with the provision that a party has to gain at least 5% of the votes nationwide in order to be represented in the Parliament.

[27] Thus, unlike the U.S. case, where hawks and doves are usually represented along bureaucratic roles, the party affiliation of the minister determines the policy direction of the Foreign Office as compared with that of the Ministry of Defense.

[28] After all, the strong state institutions of the Fifth Republic were intended to counterbal-

hile this polarization has declined considerably over the past ten years ırtly as a result of the French Left taking over the presidency),[29] the ·akness and fragmentation of the French social organizations is still iking. (Compare, for example, the ideological divisions between the merous French trade unions with the unified strength of the German deration of Trade Unions [*Deutscher Gewerkschaftsbund*].) This gen- ıl description extends into the foreign policy area.

French mass public opinion is more divided on security policy than is ıt of most other Western European nations. The divisions are strong- along party lines, which in turn reflect ideological cleavages between : Left and the Right. Concerning attitudes on defense spending, for ımple, supporters of the French Socialist Party (PSF) have been more ın 20 percentage points apart from adherents of the Gaullist Party PR), whereas the respective gap between supporters of the German :ial Democrats (SPD) and of the Christian Democrats (CDU/CSU) s only 10–15 points.[30] Moreover, the opposition to the French defense icy was never able to mobilize much societal support. Although a .jority of the French public rejected the independent nuclear force ll into the 1970s, this opposition never manifested itself in strong peace ıvements, as has frequently been the case in the U.S., West Germany, 1 Japan.[31]

Compared with France, Japan seems to be an almost homogeneous iety in which the achievement of social consensus ranks among the ,hest cultural values. Accordingly, mass public opinion on foreign pol- is among the most stable and most consistent of all industrialized nocracies.[32] Moreover, Japanese society, in particular the business nmunity, is highly organized. In the security policy area, for example, defense industry can marshal strong institutions to further its inter- ;. On the other side of the spectrum, the Japanese peace movements ıe also been well organized, particularly during the 1950s and 1960s, 1 have maintained strong links to trade unions and left-wing opposi- ı parties.[33] In 1960, for example, the renewal of the Mutual Security

: the divisions of French society. See Douglas E. Ashford, *Policy and Politics in France: ng with Uncertainty* (Philadelphia: Temple University Press, 1982).

See Paul Godt, ed., *Policy-making in France: From de Gaulle to Mitterrand* (London: er, 1989).

Data for 1979–83 are in Eichenberg (fn. 1), 186–87. For further evidence, see Eichenberg, 141, 147, 189.

See Jolyon Howorth, *France: The Politics of Peace* (London: Merlin Press, 1984).

For data, see Bobrow (fn. 4); and James A. A. Stockwin, "Japanese Public Opinion and cies on Security and Defence," in Ronald Dore and Radha Sinha, eds., *Japan and World ression* (New York: St. Martin's Press, 1987), 111–34.

For details, see Reinhold Drifte, *Arms Production in Japan* (Boulder, Colo.: Westview

Treaty with the U.S. led to the most serious political crisis in postwar Japan, as peace movements, opposition parties, and trade unions launched a massive campaign against the treaty, including the largest strikes in Japanese history.

In general, American society resembles French society more than Japanese society in terms of class divisions and ideological cleavages. Again, this has repercussions in foreign and security policy. As Eugene Wittkopf and others have shown, the bifurcation of the American people into "militant internationalists" and "cooperative internationalists" is strongly correlated with ideological divisions between conservatives and liberals and, to a lesser extent, with partisanship; the cleavages have increased since the 1970s.[34] Nevertheless, the ability of societal actors to mobilize support for their demands and to organize themselves seems to be greater in the U.S. than in France. One only has to mention the significant power of the American defense industry, as well as the importance of countervailing public interest groups lobbying for arms control and disarmament.[35]

German society seems to be less fragmented than either U.S. or French society in terms of ideological and class cleavages. Moreover, the country enjoys comparatively strong social organizations and a high level of participation in those associations. The three most important and centralized social organizations—business, trade unions, and churches—never hesitate to speak out on foreign and security issues, usually on the dovish side of the debate. The peace movements of both the 1950s and the 1980s were able to overcome their internal divisions and create strong coordinating mechanisms.[36]

POLICY NETWORKS

Concerning the policy networks, the American system seems to come closest to the society-dominated type. Constant building and rebuilding of coalitions among societal actors and political elites is fairly common

---

Press, 1986), 21–29; Kai Fabig, *Japan: Wirtschaftsriese—Rüstungszwerg* (Bochum: Studienverlag Brockmeyer, 1984); and Welfield (fn. 24), 434–41.

[34] Wittkopf (fn. 1), 34–36, 44–49. In fact, the ideological cleavages are deeper than those in the sociodemographic realm. See also Rattinger (fn. 6).

[35] On the U.S. armaments industry, see, e.g., Gordon Adams, *The Politics of Defense Contracting: The Iron Triangle* (New York: Council on Economic Priorities, 1981); Fen O. Hampson, *Un-guided Missiles: How America Buys Its Weapons* (New York: W. W. Norton, 1989); on the U.S. peace movements, see David S. Meyer, *A Winter of Discontent: The Nuclear Freeze and American Politics* (New York: Praeger, 1990); and Frances B. McCrea and Gerald E. Markle, *Minutes to Midnight: Nuclear Weapons Protest in America* (Newbury Park, Calif.: Sage, 1989).

[36] For the 1980s, see Risse-Kappen (fn. 17). For the 1950s, see Marc Cioc, *Pax Atomica: The Nuclear Defense Debate in West Germany during the Adenauer Era* (New York: Columbia University Press, 1988).

even in foreign and security policy. The openness of the political system provides the society with comparatively easy access to the decision-making process. Accordingly, corporatist structures such as the military-industrial complex linking business, military, and political interests have only a limited impact on foreign policy decisions beyond the weapons-procurement process and demands for a certain level of defense spending.[37]

The policy network in France, by contrast, seems to be almost completely dominated by the state and, above all, the presidency, particularly concerning foreign and defense policy. Even when Socialist President Mitterrand was forced into "cohabitation" with a Conservative government from 1986 to 1988, he continued to control the foreign policy-making process.[38] State dominance in the policy network is facilitated by the famous French "defense consensus" among the elites and all political parties in support of French independence in world politics and of the *force de dissuasion*. As a result, the French foreign policy-making process has been described as that of a "nuclear monarchy."[39]

Differing from the French and U.S. cases, the situation of the Germans and Japanese resembles the corporatist model. In the Federal Republic the party system not only permeates the state institutions, but it also forms the most important link between society and the political system. The two major parties—the Christian Democratic Union (CDU) and the Social Democratic Party (SPD)—are essentially catchall organizations that integrate rather divergent societal demands. Sometimes intraparty divisions are greater than the cleavages between parties, as was the case with the CDU and *Ostpolitik* during the 1970s and with the SPD and deterrence during the 1980s. Additionally, institutional arrangements as well as the political culture emphasize consensus building and the mutually beneficial reconciliation of diverging societal interests (*Interessenausgleich*). There is a clear analogy between the domestic emphasis on social partnership and a foreign policy trying to achieve its goals by security partnership with Germany's neighbors.[40]

Consensus building is of similar and probably even greater importance

[37] This, at least, is the finding of various studies. See, for example, Stephen Rosen, ed., *Testing the Theory of the Military Industrial Complex* (Lexington, Mass.: Lexington Books, 1973); Krell (fn. 1); and Müller (fn. 1). On the military procurement process, see Hampson (fn. 35).

[38] See Phillippe G. LePrestre, "Lessons of Cohabitation," in LePrestre, ed., *French Security Policy in a Disarming World* (Boulder, Colo.: Lynne Rienner, 1989), 15–47.

[39] See Philippe G. LePrestre, "Lessons of Cohabitation," in LePrestre, ed., *French Security* Howorth and Patricia Chilton, eds., *Defence and Dissent in Contemporary France* (London: Croom Helm, 1984).

[40] A content analysis of party and government documents for 1977–84 showed that "international cooperation" was by far the most frequently mentioned foreign policy objective. See Risse-Kappen (fn. 17).

in Japan. The need to bring almost every relevant player into the deci sion-making process and to strike a balance between different views i frequently cited as a major reason that Japanese foreign policy seems s slow moving and low-key.[41] However, to portray the Japanese polic network as "corporatism without labor"—that is, having strong and in timate linkages between the state bureaucracy, the LDP, and bi business[42]—does not seem to make as much sense in security policy. A military-industrial complex that includes the Defense Agency and th weapons industry, while having increased its role during the defens buildup of the 1980s,[43] by no means dominates Japanese security policy Even more than in the West German case, the legacy of World War I prevented the rise of a new Japanese militarism. A firm national consen sus including all relevant elite factions, the opposition, and society insure that Japanese security policy has been first and foremost a matter of for eign economic policy. As a result, military policy seems to be one issue-area in Japanese politics that does not systematically exclude labor and the Left from the policy-making process, at least not since the 1960 Security Treaty crisis.

Table 1 summarizes the similarities and differences in the domestic structures of the four liberal democracies under consideration here.

TABLE 1
DOMESTIC STRUCTURES

	U.S.A.	*W. Germany*	*Japan*	*France*
Political system	decentralized	intermediate level of centralization	intermediate level of centralization	centralized
Society	heterogeneous, weak organizations	heterogeneous, strong organizations	homogeneous, strong organizations	heterogeneous, weak organizations
Policy network	society-dominated	democratic, corporatist	quasi-corporatist	state-dominated

[41] On the importance of consensus building in Japanese politics and particularly in foreign policy, see Takeshi Ishida, *Japanese Political Culture* (New Brunswick, N.J.: Transaction Books, 1983), 3–22; Douglas Mendel, "Public Views of the Japanese Defense System," in Buck (fn. 23), 150; and Scalapino (fn. 23), xv–xviii.

[42] See, for example, T. J. Pempel, *Politics and Policy in Japan: Creative Conservatism* (Philadelphia: Temple University Press, 1982). For a critique of this analysis evaluating public policy in Japan as a more open process, see Kent Calder, *Crisis and Compensation: Public Policy and Political Stability in Japan, 1949–1986* (Princeton: Princeton University Press, 1988).

[43] See Drifte (fn. 33).

The comparison leads to several assumptions about the impact of public opinion on foreign policy in the four countries. The comparatively open and decentralized American political system and its society-dominated policy network should provide public opinion with ample opportunity to affect policy outcomes. In short, the interaction of public opinion and elites in the decision-making process is expected to come comparatively close to the bottom-up model. In contrast to the U.S. case, French public opinion is likely to play only a marginal role in the foreign policy–making process. The Fifth Republic's centralized political system and a fragmented societal structure make it difficult to build a public consensus on policy issues and would be expected to limit the public impact on foreign and security policy. The German domestic structure, in which comparatively strong state institutions deal with well-organized societal actors in a democratic corporatist network, suggests that public opinion influences foreign policy mainly through the party system and that interactive patterns prevail between elites and masses. Finally, the Japanese case is expected to resemble Germany insofar as both countries possess strong political institutions and corporatist policy networks. There is a difference, however; Japanese society is more homogeneous than German society, and the public has been less divided on foreign and security policy issues.

Withering Threat: Public Opinion Reactions to Soviet Foreign Policy, 1980–90

To isolate the role of domestic structures as the intervening variable between public opinion and policy decisions requires a case in which the influence of the international environment appears constant for the four countries and in which under ideal conditions public attitudes are also more or less similar. From the late 1970s to the late 1980s the four countries were exposed to a drastically changing Soviet policy. Moscow's conventional and nuclear arms buildup of the 1970s was followed by the intervention in Afghanistan and the stalemate in superpower arms control during the early 1980s. From 1985–86 on the new Soviet leader, Mikhail Gorbachev, initiated the politics of *perestroika*, which resulted in a turnaround of Soviet security policy toward "common security." Moscow accepted the zero option for intermediate-range nuclear forces (INF), withdrew its troops from Afghanistan, announced unilateral troop cuts in Europe, and finally revoked the Brezhnev doctrine of limited sovereignty for Eastern Europe, thereby triggering the democratic revolutions of 1989. How did mass public

opinion in the four countries perceive these drastic changes in Sovie foreign policy?[44]

Figure 1 reveals that, with the exception of Japan, the public percep tion of a Soviet threat declined from unusually high levels in the earl 1980s to its virtual disappearance in the late 1980s. The initial publi reaction to Soviet actions in the late 1970s and early 1980s, particularl the intervention in Afghanistan, was unanimously hostile and negative While the most dramatic change from a very high degree of threat perception occurred in the U.S., the West Germans were the first to reac positively to the turnaround of Soviet foreign policy. In 1986, that is, during the very first year of Gorbachev's "peace offensive" and one year *before* the INF treaty (which particularly affected German security) was concluded, a majority became convinced that the Soviet threat had disappeared. The threat as perceived by the French declined in 1987–88, whereas it was only in late 1988 that more than 50 percent of Americans thought that the Soviet threat had withered away. Finally, the "Gorbachev revolution" apparently did not affect Japanese public opinion. Not only did comparatively more Japanese perceive a Soviet threat throughout the 1980s, but the percentage remained high even in 1987–88.

This evaluation is confirmed by an opinion poll taken immediately after the Washington summit in December 1987 at which the INF treaty was signed. Seventy-three percent of the Germans, 55 percent of the Americans, 54 percent of the French, but only 34 percent of the Japanese thought that the Soviet Union was becoming a more trustworthy nation as a result of the reforms initiated by Gorbachev.[45] Thus, the Germans appeared to be the most enthusiastic supporters of Gorbachev, while the Japanese remained unaffected by the new Soviet foreign policy, and the Americans and the French were in between.

How did the changes in threat perceptions translate into attitudes on how to deal with the Soviet Union? The data reveal a major difference between public opinion trends in the United States as compared with those in France, Germany, and Japan. Only in the U.S. did the negative reaction to Moscow's policies during the late Brezhnev era result in in-

[44] The following data are taken from a variety of opinion polls. Although I have tried to use as many comparable data as possible, the wording of questions sometimes differed considerably from country to country. As a result, a note of caution is appropriate regarding the interpretation of these data. For example, I will refrain from analyzing minor changes in percentage. Also, I can only present here data on the aggregate level of overall mass public opinion. For comprehensive studies with breakdowns according to age, education, and political orientation, see Eichenberg (fn. 1); and Wittkopf (fn. 1).

[45] Data in Hastings and Hastings, eds., *Index to International Public Opinion, 1987–1988* (New York: Greenwood, 1989), 574.

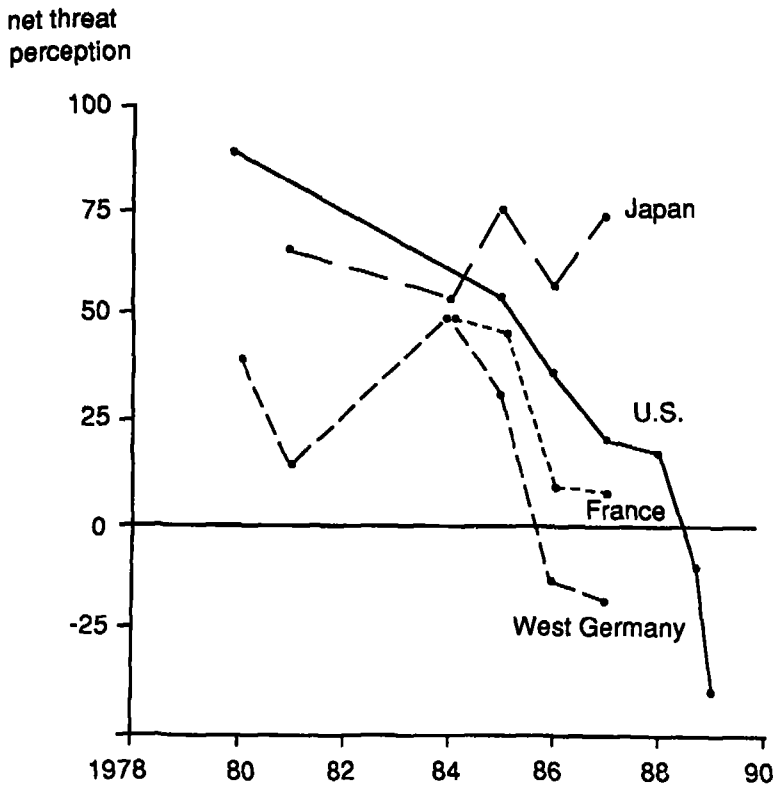



FIGURE 1
NET THREAT PERCEPTION OF THE USSR[a]

SOURCES: U.S.: *Americans Talk Security*, no. 12 (1989), 58; CBS/*New York Times* polls; France: USIA polls; Michel Girard, "L'opinion publique et la politique extérieure," *Pouvoirs*, no. 51 (1989); Szabo (fn. 11); West Germany: Hans Rattinger, "The INF Agreement and Public Opinion in West Germany" (Manuscript, March 1988); Gebhard Schweigler, *Grundlagen der aussenpolitischen Orientierung der Bundesrepublik Deutschland* (Baden-Baden: Nomos, 1985), 83; Szabo (fn. 11); Japan: Hastings and Hastings, *Index to International Public Opinion, 1981–1982ff.* (New York: Greenwood, 1983ff.).

[a] "Net threat perception" is defined as the percentage of those perceiving *a Soviet threat* minus the percentage of those perceiving *no threat*.

creasing support for hawkish policies. From late 1979 to early 1981 a majority of Americans favored increases in defense spending (see Figure 2) and opposed the ratification of the SALT II treaty. However, the support for a tougher stance disappeared as quickly as it had emerged. By mid-1981, 65 percent of Americans favored the resumption of nuclear arms talks. By 1982 support for increased defense spending was back at the

% net support
40
20
0
-20
-40
Fran
Japa
U.S.
West
Gerr
1976 78 80 82 84 86 88

FIGURE 2
NET SUPPORT FOR DEFENSE SPENDING[a]

SOURCES: U.S.: CBS/*New York Times* polls; Sean M. Lynn-Jones, "Lulling and Stimulating Effects of Arms Control," in A. Carnesale and R. Hass, eds., *Superpower Arms Control* (Cambridge, Mass.: Ballinger, 1987), 237; France and West Germany: Eichenberg (fn. 1), 161–62; Japan: Hastings and Hastings, *Index to International Public Opinion, 1979–1980ff.* (New York: Greenwood, 1981ff.).

[a] "Net support for defense spending" is defined as the percentage of those favoring *increases* in defense spending minus the percentage of those supporting defense *reductions*. The graph plots change rather than continuity, since it does not include those who favored "leaving the defense budget as it is."

1978 level; at the same time, a near consensus emerged in the U.S. mass public regarding a bilateral freeze of the superpowers' nuclear arsenals.[46] It should be noted that these attitude changes occurred *before* Gorbachev assumed power. They might be explicable by the "politics of opposites" as described by Miroslav Nincic, who suggested that U.S. policies that

[46] Data on support for nuclear arms control are taken from various U.S. polls on file at the Yale Public Opinion Research Project, New Haven.

are either too dovish or too hawkish toward the Soviet Union would be opposed by the public mainstream.[47]

Compared with the shifts in U.S. public opinion, French, German, and Japanese public opinion has been far more stable in support of or in opposition to specific policies. This is most obvious in the attitudes toward defense spending, which remained largely unaffected by the changes in Soviet policy (see Figure 2). From the late 1970s to the mid-1980s a majority of Germans preferred a stable defense budget; the number favoring reductions increased in the late 1980s. The same applies roughly for the French and the Japanese, although comparatively more Japanese than Germans advocated an increased defense budget. In sum, however, the perceived Soviet threat prior to Gorbachev did not result in demands for a military buildup in the three countries.

If anything, the Japanese were more hawkish than the French and the Germans during the early 1980s. In 1981, 31 percent maintained that a military balance was the most important source of security, while only 12 percent and 8 percent believed in détente and arms control, respectively. Two years later, however, Japanese public opinion was back in line with the French and West Germans, who had always thought that détente and/or arms control were more important than maintaining a military balance with the Soviet Union.[48] A particularly interesting finding is revealed by a 1984 Japanese poll. Fifty-nine percent of those who perceived the USSR as a threat nevertheless favored the improvement of peaceful diplomatic relations between the two countries, as compared with only 21 percent who wanted to strengthen military security.[49] In sum, the public consensus in Japan on basic foreign policy problems emphasizing economic strength as well as peaceful diplomacy remained largely intact throughout the 1980s.

West German public opinion came out as the most dovish throughout the decade. In sharp contrast to the Americans, 74 percent of the Germans continued to support détente and *Ostpolitik* in the midst of the Afghanistan crisis in early 1980. Nuclear-arms-control efforts such as the zero option for INF enjoyed near consensus in the 70–90 percent range.[50]

[47] See Nincic, "The United States, the Soviet Union, and the Politics of Opposites," *World Politics* 40 (July 1988), 452–75; and idem, "U.S. Soviet Policy and the Electoral Connection," *World Politics* 42 (April 1990), 370–96. However, the "politics of opposites" seems to be a unique U.S. phenomenon. A similar cycle is not observable in the three other countries.

[48] See data in Gregory Flynn, *Public Images of Western Security* (Paris: Atlantic Institute for International Affairs, 1985), 65.

[49] Data in Hastings and Hastings, eds., *Index to International Public Opinion, 1984–1985* (New York: Greenwood, 1986), 237.

[50] For data on *Ostpolitik*, see Elisabeth Noelle-Neuman and Edgar Piel, eds., *Allensbacher*

FIGURE 2
NET SUPPORT FOR DEFENSE SPENDING[a]

SOURCES: U.S.: CBS/*New York Times* polls; Sean M. Lynn-Jones, "Lulling and Stimulating Effects of Arms Control," in A. Carnesale and R. Hass, eds., *Superpower Arms Control* (Cambridge, Mass.: Ballinger, 1987), 237; France and West Germany: Eichenberg (fn. 1), 161–62; Japan: Hastings and Hastings, *Index to International Public Opinion, 1979–1980ff.* (New York: Greenwood, 1981ff.).

[a] "Net support for defense spending" is defined as the percentage of those favoring *increases* in defense spending minus the percentage of those supporting defense *reductions*. The graph plots change rather than continuity, since it does not include those who favored "leaving the defense budget as it is."

1978 level; at the same time, a near consensus emerged in the U.S. mass public regarding a bilateral freeze of the superpowers' nuclear arsenals.[46] It should be noted that these attitude changes occurred *before* Gorbachev assumed power. They might be explicable by the "politics of opposites" as described by Miroslav Nincic, who suggested that U.S. policies that

[46] Data on support for nuclear arms control are taken from various U.S. polls on file at the Yale Public Opinion Research Project, New Haven.

are either too dovish or too hawkish toward the Soviet Union would be opposed by the public mainstream.[47]

Compared with the shifts in U.S. public opinion, French, German, and Japanese public opinion has been far more stable in support of or in opposition to specific policies. This is most obvious in the attitudes toward defense spending, which remained largely unaffected by the changes in Soviet policy (see Figure 2). From the late 1970s to the mid-1980s a majority of Germans preferred a stable defense budget; the number favoring reductions increased in the late 1980s. The same applies roughly for the French and the Japanese, although comparatively more Japanese than Germans advocated an increased defense budget. In sum, however, the perceived Soviet threat prior to Gorbachev did not result in demands for a military buildup in the three countries.

If anything, the Japanese were more hawkish than the French and the Germans during the early 1980s. In 1981, 31 percent maintained that a military balance was the most important source of security, while only 12 percent and 8 percent believed in détente and arms control, respectively. Two years later, however, Japanese public opinion was back in line with the French and West Germans, who had always thought that détente and/or arms control were more important than maintaining a military balance with the Soviet Union.[48] A particularly interesting finding is revealed by a 1984 Japanese poll. Fifty-nine percent of those who perceived the USSR as a threat nevertheless favored the improvement of peaceful diplomatic relations between the two countries, as compared with only 21 percent who wanted to strengthen military security.[49] In sum, the public consensus in Japan on basic foreign policy problems emphasizing economic strength as well as peaceful diplomacy remained largely intact throughout the 1980s.

West German public opinion came out as the most dovish throughout the decade. In sharp contrast to the Americans, 74 percent of the Germans continued to support détente and *Ostpolitik* in the midst of the Afghanistan crisis in early 1980. Nuclear-arms-control efforts such as the zero option for INF enjoyed near consensus in the 70–90 percent range.[50]

[47] See Nincic, "The United States, the Soviet Union, and the Politics of Opposites," *World Politics* 40 (July 1988), 452–75; and idem, "U.S. Soviet Policy and the Electoral Connection," *World Politics* 42 (April 1990), 370–96. However, the "politics of opposites" seems to be a unique U.S. phenomenon. A similar cycle is not observable in the three other countries.

[48] See data in Gregory Flynn, *Public Images of Western Security* (Paris: Atlantic Institute for International Affairs, 1985), 65.

[49] Data in Hastings and Hastings, eds., *Index to International Public Opinion, 1984–1985* (New York: Greenwood, 1986), 237.

[50] For data on *Ostpolitik*, see Elisabeth Noelle-Neuman and Edgar Piel, eds., *Allensbacher*

Moreover, an increasing number of Germans favored unilateral Weste arms reductions to encourage Soviet disarmament; by 1987 more than percent supported such a strategy of reassurance.[51]

A March 1985 international opinion poll asking for people's preferre Western strategy toward the new Soviet leadership confirms the overa pattern (see Figure 3). Once again, there is a striking contrast betwee German, French, and Japanese public opinion on the one hand and th of the U.S. on the other. Whereas only a minority in each of the fou countries supports hawkish policies emphasizing military strength, plurality of Americans seems to be more cautious than the dovis Europeans, favoring instead an "owlish" approach (that is, combinin military strength with the preparedness to compromise). Again, Wes German public opinion is most supportive of arms control and détente.

While it is not the purpose of this paper to explain the patterns o public opinion, a note is offered on the developments in public opinio in the U.S. as compared with developments in France, Germany, an Japan. The data are consistent with an explanation according to whic in Western Europe (and probably in Japan, too) negative feelings towar the Soviet Union did not determine the preferences for specific securit policies; rather, the fear of war and of the arms race in general as well as the perception of U.S. policies were of at least equal significance.[52] Indeed, during the early 1980s a plurality of West Germans and French perceived U.S. policies as risking war rather than promoting peace, and the general fear of war reached unusual highs.[53] In sum, even before the cold war was actually over, large segments of Western European—and, maybe, Japanese—public opinion subscribed to a "post–cold war" international environment in which the Soviet threat no longer dominated the agenda.[54]

To conclude, the data reveal similarities as well as differences in the public attitudes of the four countries toward the Soviet Union. While the

Jahrbuch der Demoskopie 1978–1983 (Munich: Saur, 1983), 637; on nuclear arms control, see various USIA polls, quoted in Szabo (fn. 11).

[51] Data in Elisabeth Noelle-Neumann, "Wenn das Gefühl der Bedrohung schwindet," *Frankfurter Allgemeine Zeitung*, July 22, 1988. In comparison, about 80 percent of the American public rejected such unilateral steps during the 1980s; data according to Roper and other U.S. polls, on file at the Yale Public Opinion Research Project, New Haven.

[52] For this explanation, see Don Munton, "NATO up against the Wall: Changing Security Attitudes in Germany, Britain, and Canada" (Paper presented at the Annual Meeting of the *International Studies Association*, Washington, D.C., April 1990).

[53] For data regarding the perception of U.S. policies, see Szabo (fn. 11).

[54] This consideration would also explain why Nincic's "politics of opposites" (fn. 47, 1988) was not observable in Western Europe and Japan. The model is based on a pattern of attitudes in which the perception of Soviet foreign policy still dominates the preferences for specific security policies.

Question: *In view of the change in Soviet leadership, in which one of the ways listed below should the West proceed in its negotiation with the Soviet Union in order to reach an agreement on arms control?*

—"tough/no compromise + military buildup" (Hawks)

—"more moderate + maintain current military strength" (Owls)

—"offer compromise to encourage the Soviets" (Doves)

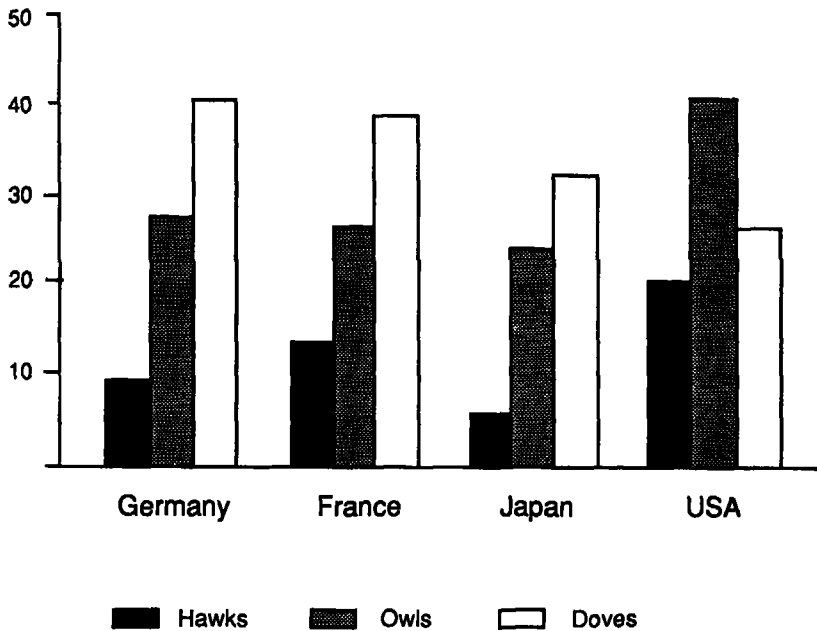



FIGURE 3
POLICY TOWARD THE SOVIET UNION

SOURCE: Hastings and Hastings, *Index to International Public Opinion, 1985–1986* (New York: Greenwood, 1987), 591.

change in the perception of threat occurred at a different time for each, by 1988 the Soviet threat had withered away everywhere except Japan. Moreover, West Germans, French, and Japanese continuously preferred détente and arms control policies to a military buildup, albeit to different degrees. The deviant case was public opinion in the U.S., which favored rather hawkish policies during the early 1980s. But even there arms control was supported again as early as 1982–83. Thus, there have been discernible differences in public attitudes across the four countries. But the similarities seem to be substantial enough to merit a comparative study of their policy impact.

The Public Impact on Foreign Policy toward the Soviet Union during the 1980s

A comparison of the above data on policy preferences toward the Sovie Union (see Figure 3) with the actual policies of the four countries indicates some preliminary results. On the one hand, Reagan's initial, owlisl policy toward Gorbachev as well as the West German embrace of th new Soviet foreign policy right from the beginning—"Genscherism"—seem to be roughly in line with the public opinion trends in the twc countries. On the other hand, both the French and the Japanese conducted rather cautious and sometimes even hostile policies toward Gorbachev, policies that seem to have been out of touch with public opinion, particularly in the case of France. This pattern—of a match of public opinion and policy in the cases of the U.S. and the Federal Republic of Germany, less of a match in Japan, and no match in France—is exactly what one would expect using the domestic structure approach. However, detailed process tracing is required to confirm the proposition.

TWO SIDES OF OPENNESS: THE UNITED STATES

Given the domestic structure of the United States, the interaction of public opinion, societal forces, and political elites in the foreign policy process would be expected to resemble the bottom-up model. Indeed, a longitudinal analysis of public opinion and policy changes from 1935 to 1979 revealed a congruence between public opinion changes and policy shifts, especially in cases of high issue salience and drastic changes in public attitudes. Benjamin Page and Robert Shapiro found essentially no difference between domestic and foreign policy issues. More important, in at least half of the cases the policy changes *followed* the shifts in public opinion.[55]

U.S. policy toward the Soviet Union from the late 1970s to the late 1980s reveals the general validity of the hypothesis. Right after the conclusion of the SALT I treaties, the American security elite became increasingly split on the virtues of arms control and détente. During the late 1970s the U.S. public was presented with two different approaches for dealing with the Soviet Union: the views of the traditional arms-control establishment represented in the Carter administration and an outlook

[55] See Page and Shapiro (fn. 3). Note, however, that the study does not prove the bottom-up model. It correlates public opinion data with policy outcomes and does not control for the impact of elite opinion. It would be consistent with the data to assume, for example, that elite opinion changed first and then affected the general public, which in turn led to policy changes.

most prominently expressed by the Committee on the Present Danger.[56] As the data presented above reveal, the change in public opinion toward "peace through strength" policies took place in late 1979/early 1980. The Soviet intervention in Afghanistan and not just the conservative part of the U.S. elites tipped the balance in favor of the antidétente mood among the public. This contributed to the election victory of Ronald Reagan, and, as a result, to the anti-SALT coalition.[57] In other words, while the events of the late 1970s and early 1980s do not confirm a simplistic version of the bottom-up model, they nevertheless show that U.S. public opinion played a role in bringing the anti-SALT coalition to power.

A similar process can be observed in 1982–84. Reagan's military buildup and his cold war rhetoric quickly eroded the public support for his defense policy. The "politics of opposites" (Nincic) worked (see data above). This change as well as the stable public consensus against the first use of nuclear weapons provided issue publics such as the freeze campaign with a window of opportunity. The nuclear protest was a bottom-up movement that originated outside of Washington's arms-control elite.[58] It quickly gained ground, penetrated the more skeptical arms-control community, and found support among numerous social organizations such as the Catholic church. Given the openness of the U.S. political system, the movement found easy access to the political elites. In 1983 the House of Representatives adopted a freeze resolution. Congress became increasingly active on issues of nuclear arms control.

As a result of these coalition-building processes in the policy network, the power balance in the Reagan administration between hardliners in the Pentagon and more pragmatic conservatives in the State Department slowly shifted in favor of the latter.[59] By 1984 and continuing throughout 1985, that is, *before* Mikhail Gorbachev entered upon the scene, Ronald Reagan had softened his rhetoric and adopted a more compromising stance on arms control. The superpower relationship improved, nuclear-

[56] On the latter, see Max Kampelmann, ed., *Alerting America: The Papers of the Committee on the Present Danger* (Washington, D.C.: Pergamon-Brassey's, 1984); and Jerry W. Sanders, *Peddlers of Crisis: The Committee on the Present Danger* (Boston: South End Press, 1983).

[57] For details, see Bernd W. Kubbig, *Amerikanische Rüstungskontrollpolitik: Die innergesellschaftlichen Kräfteverhältnisse in der ersten Amtszeit Reagans (1981–1985)* (Frankfurt am Main: Campus, 1988); and Barry Posen and Stephen Van Evera, "Defense Policy and the Reagan Administration: Departure from Containment," *International Security* 8 (Summer 1983), 3–45. The best account of U.S. nuclear arms control policy during Reagan's first term is Strobe Talbot, *Deadly Gambits* (New York: Knopf, 1984).

[58] The most comprehensive account of the freeze movement is Meyer (fn. 35).

[59] For details, see Strobe Talbot, *The Master of the Game: Paul Nitze and the Nuclear Arms Race* (New York: Knopf, 1988). See also Ernst-Otto Czempiel, "U.S. Policy towards the Soviet Union under Carter, Reagan, and Bush" (Paper presented to the Annual Convention of the International Studies Association, Washington, D.C., April 10–14, 1990); and Joseph (fn. 9), chap. 5.

arms-control talks resumed in early 1985, and the U.S. defense bud[
began to decline by 1986. However, the freeze campaign failed to achi(
its immediate goals, and the shift in Reagan's policy did not prodı
substantial arms-control results until the late 1980s (INF Treaty 198
But public opinion clearly affected the ability of the issue public a:
interest groups to promote their cause and thereby influence the coa
tion-building process among the elites, which in turn produced tangib
albeit limited, results in U.S. policies.

The impact of public opinion on the conduct of U.S. policy towa
the Soviet Union was also noticeable in the spring of 1989. When tl
Bush administration came into power, government officials conducted
comprehensive review of Reagan's policies and apparently conclud(
that a more cautious approach should be adopted toward Moscow. Hov
ever, as discussed above, by that time the Soviet threat had already witl
ered away for the American public, so much so that the public reactc
negatively to Bush's early foreign policy statements. The administratio
quickly adjusted to the public mood, however, and in June 1989 the pre:
ident assured the public and the allies that the U.S. was firmly supporti\
of Gorbachev and the politics of perestroika.[60]

While public opinion did not determine U.S. policy toward the Sovi(
Union during the 1980s, it nevertheless left a discernible mark on th
decision-making process. The analysis confirms that the U.S. domesti
structure is a comparatively open system that allows societal actors t(
mobilize support and to affect the balance of forces within the polic
network. Nevertheless, the fragmentation and decentralization of th(
political system work against the stabilization and institutionalization o
policies. Policymakers were able to decrease the salience of the issues ir
question by symbolically adjusting to the public mood, as was the case ir
1984–85. The public impact was more limited regarding the substance ol
policies, mainly because the domestic structure does not provide institu-
tional support for lasting consensus among the elites (as, for example, a
strong party system would). In sum, the very openness of the U.S. system
to societal pressures also limits their impact.

WEAKNESSES OF A "NUCLEAR MONARCHY": FRANCE

The French domestic structure, as compared with that of the U.S., provides public opinion with only limited influence on foreign policy deci-

[60] This account is based largely on newspaper articles. See, e.g., "Cheney Remarks on Soviet Future Ruffle the White House's Feathers," *New York Times*, May 1, 1989; "Bush Asks an End to Divided Europe," *New York Times*, June 1, 1989; and "Bush Policy Makers Reach Uneasy Balance on an Approach to the Soviets," *New York Times*, July 2, 1989.

With the possible exception of the ending of the Algerian war,[61] is indeed not much evidence that any of the major foreign policy ons by the presidents of the Fifth Republic were taken in response olic opinion. This holds true for the building of an independent ır force, the withdrawal from NATO's military command, the policy ente toward the Soviet Union in the 1960s and 1970s, and the gradift toward more pro-European attitudes under de Gaulle's succesn most cases public opinion was deeply split on the issues. Morehe French presidents were usually able to create a public consensus rting their foreign policies *after* the event.[62]

imilar pattern prevailed during the 1980s. As shown above, French opinion did not develop in a markedly more hawkish way than ıblic opinion in Germany or Japan. Nevertheless, the new Socialist ent, Mitterrand, adopted a more hostile approach toward the Sonion than that held by his predecessors, who had basically contine legacy of de Gaulle's independent policy of détente. While Mitd's policy change did not reflect trends in mass public opinion, it line with parts of the issue public and the political elites. French ing intellectuals who had been fairly pro-Soviet during the 1950s 960s "discovered" repression in the USSR in the early 1970s and l overwhelmingly anticommunist as a result. When President Gis)'Estaing tried to continue détente in the aftermath of the Soviet ention in Afghanistan, he was chastised by the intellectuals and the . His successor, Mitterrand, changed the policy, not least of all for tic reasons, in an effort to isolate and marginalize the French unist Party (PCF) on the Left.[63]

limited role of French society and public opinion in influencing decisions was also apparent during the early 1980s when France e only major Western European country that did *not* face a signifpublic security debate and protest movements against nuclear ns. This may be explained in part by the fact that the new U.S. issiles were not about to be deployed in France. Still, the French d nuclear arms control and were as opposed to Reagan's policy as y other European public. However, the French peace movement developed into a mass movement; on the contrary, it was not only

Grosser (fn. 22), 161–72.

details, see Cohen (fn. 22); Grosser (fn. 22); Howorth (fn. 31); and Aldrich and (fn. 21).

details, see Cohen (fn. 22); Julius W. Friend, *Seven Years in France* (Boulder, Colo.: w Press, 1989), 75–81; and Michel Tatu, "Valéry Giscard d'Estaing et la détente," in ohen and Marie Claude Smouts, eds., *La politique extérieure de Valéry Giscard d'Esta*-s: Presses de la fondation nationale des sciences politiques, 1985), 196–217.

marginalized by the mass media but also internally split along ideologi
lines from the very beginning. In that sense, it reflected the fragmen
tion of French society.[64]

While the French centralized political institutions and the sta
dominated policy network seem to constrain the public impact on f
eign policy, there are a few instances in which public opinion apparen
did play a role. One recent example was Mitterrand's decision in 1987
endorse the INF treaty, despite the objections of his conservative forei
and defense ministers and of the majority of the French security coı
munity, who thought that the zero option was detrimental to Weste
deterrence policy. One year before the presidential elections, Mitterra
was apparently aware that French public attitudes overwhelmingly e
dorsed the U.S.-Soviet agreement. The decision is an example of pre
dential dominance in foreign policy matters even in times of "cohabit
tion"; it also suggests that public opinion sometimes affects Fren
foreign policy directly, through presidential decisions rather th
through coalition-building processes among the elites.[65]

While French public opinion only marginally influenced foreign pc
icy decisions during the 1980s, the process did not resemble the top-dow
model. The French public did not just follow the policies of the elites
of Mitterrand. A comparison between public opinion and elite opinic
shows that the attitude gap between elites and masses in France is one
the largest in Europe.[66] French public opinion seems simply to be discoı
nected from the policy-making process.

The limited role of French society in the country's foreign policy
making process does not, however, result in a highly active policy uı
hampered by domestic constraints and able to concentrate solely on e
ternal factors, as the concept of "state strength" would assume. The pow
erful, centralized state institutions of the Fifth Republic were designe
to integrate a heavily divided and fragmented society. The foreign polic
of de Gaulle and, albeit to a lesser extent, of his successors was intimatel
linked to the efforts to strengthen the societal support basis for the insti

[64] In addition, there were historical reasons for the isolation of the French peace move
ment. The notion of pacifism was identified with the politics of appeasement in the 1930s
and the largest French peace movement, the Mouvement de la Paix, was never able to dis
tance itself convincingly from the Communist Party. For details, see Howorth and Chilton
(fn. 39).

[65] In this particular case, Mitterrand was apparently influenced by public opinion not onl
in France but also in West Germany and by the fear that rejection of zero INF would lead to
German neutralism. Cf. Jolyon Howorth, "Consensus and Mythology: Security Alternative
in Post-Gaullist France," in Aldrich and Connell (fn. 21), 16–34; and LePrestre (fn. 38).

[66] See Eichenberg (fn. 1), 222–24.

tutions of the Fifth Republic.[67] To this day the official ideology emphasizes French unity, *grandeur*, and independence from external influences and alliances, notions that cut across ideological cleavages.[68] As a result, the famous French consensus on foreign and defense affairs remains fragile. The French state pays a price for excluding society from the foreign policy–making process. Rather than being able to pursue an autonomous and active policy. French decision makers seem to be constantly concerned with stabilizing a vulnerable elite consensus.

FROM TOP-DOWN TO BOTTOM-UP: THE FEDERAL REPUBLIC OF GERMANY

A review of West German foreign policy reveals that the impact of public opinion increased over time and that the decision-making process became more democratic. In the 1950s, when Chancellor Adenauer integrated the Federal Republic into the Western alliance and the European Community, general public opinion was as deeply divided on the issue as were the political elites and the parties. However, Adenauer went ahead, and by about 1960 the public as well as the opposition had accepted his security policies.[69]

This top-down pattern gradually changed over the next two decades. In 1968–70, when Chancellor Brandt began his *Ostpolitik* and concluded the treaties with the Soviet Union and Eastern Europe, public opinion as well as the political elites were again divided on the policy. The Christian Democratic opposition (CDU) launched a vehement campaign against *Ostpolitik* and was even successful in depriving the government of its majority in Parliament. This time, however, and unlike the 1950s, the populace decided the issue: Brandt won the 1972 national elections primarily on the basis of *Ostpolitik*.[70] Throughout the 1970s the public consensus in favor of détente stabilized, and by 1980, despite the Soviet arms

[67] See Philip G. Cerny, "Gaullism, Nuclear Weapons, and the State," in Howorth and Chilton (fn. 39), 46–74; and David Hanley, "The Parties and the Nuclear Consensus," ibid., 75–93. On the institutional uncertainty of the French system as a dominant feature of the country's policy network, see Ashford (fn. 28).

[68] "Independence," for example, was understood in an anti-Soviet sense by the French conservatives, in an anti-American way by the Left, and in a neutralist sense by nationalists at both ends of the political spectrum.

[69] For details, see Arnulf Baring, *Am Anfang war Adenauer: Die Entstehung der Kanzlerdemokratie* (Munich: Deutscher Taschenbuch-Verlag, 1971); Cioc (fn. 36); and Hans-Karl Rupp, *Außerparlamentarische Opposition in der Ära Adenauer* (Cologne: Pahl-Rugenstein, 1970).

[70] For details, see Helga Haftendorn, *Sicherheit und Entspannung: Zur Außenpolitik der Bundesrepublik Deutschland, 1955–1982* (Baden-Baden: Nomos, 1983); Wolfram F. Hanrieder, *Germany, America, Europe* (New Haven: Yale University Press, 1989), 170–219; Clay Clemens, *Reluctant Realists: The Christian Democrats and West German Ostpolitik* (Durham, N.C.: Duke University Press, 1989); and Christian Hacke, *Die Ost- und Deutschlandpolitik der CDU/CSU* (Cologne: Wissenschaft und Politik, 1975).

buildup and the intervention in Afghanistan, it was firmly in p shown in the data above.

The opposition Christian Democrats, although internally di continued to oppose détente and in 1980 conducted another Ostpolitik election campaign, hoping it would succeed because of th terioration in East-West relations. Unlike the anti-SALT coalition in U.S., however, they failed. In November 1980, that is, two years be he became chancellor, Helmut Kohl finally declared the battle over, a the CDU accepted Ostpolitik.[71] In sum, general public opinion was cruc in affecting the coalition-building process within the party system ar thus, in bringing about the German détente consensus. Unlike t French case, the domestic structure of the Federal Republic not or proved to be more open to societal influences, but it also institutionaliz the consensus in the policy network, thus allowing for a highly acti *Ostpolitik*.

General public opinion and societal actors also made a difference du ing the nuclear debate of the 1980s.[72] The nuclear controversy original began at the elite level—mainly within the SPD—in 1977–78, in tl debate over the neutron bomb. NATO's INF decision of 1979, the ne "cold war" between the U.S. and the USSR, and the election of Ronal Reagan then triggered mass opposition. New peace movements launche their protests against the deployment of INF missiles. As with the free campaign in the U.S., the protesters in Germany relied on general publ opinion as a support basis regarding two issues: (1) opposition to a bellig erent U.S. policy and (2) opposition to nuclear weapons as instrumen of war fighting and to their deployment on German soil.

While the peace movements failed to prevent the INF deployment i 1983, they had a profound impact on the coalition-building proces among the elites. First, support for vigorous arms-control efforts in creased within all parties, in particular within the CDU. Second, th "disarmers" entered the party system. Since 1983 policy decisions withi the SPD elite have resulted from policy compromises between disarmer and the traditional arms-control establishment, but excluding member of the right wing of the party, such as former Chancellor Helmu Schmidt. Moreover, a disarmament party, the Greens, entered the polit ical scene. Third, the nuclear elite consensus broke down as a result. By

[71] For details, see Risse-Kappen (fn. 17), pt. B.

[72] For the following, see details ibid. On the peace movements in particular, see Josef Janning et al., eds., *Friedensbewegungen* (Cologne: Wissenschaft und Politik, 1987); and Thomas Rochon, *The Politics of the Peace Movement in Western Europe* (Princeton: Princeton University Press, 1988).

4 the SPD had adopted an antinuclear stance. In sum, the coalitions thin the policy network gradually moved to the Left—in line with ends in general public opinion and in response to the protests of the sue publics.

The consequences of both the détente debate and the antinuclear protests were seen in the mid- to late 1980s. Germans were the first to embrace Gorbachev's new policy, because it responded favorably to what had previously emerged as a consensus on *Ostpolitik* and common security. Moreover, the opposition to new nuclear weapons increasingly included the CDU, the consequences of which became apparent during NATO's debate on the modernization of battlefield nuclear weapons in the spring of 1989. There was a coincidence of public opinion, left-wing antinuclearism, and conservative fears that short-range systems would "singularize" Germany in the aftermath of the INF treaty. As a result, Chancellor Kohl had no choice but to resist a modernization decision by NATO.

To conclude, the role of public opinion in West German security policy changed over time and moved closer to the bottom-up model during the 1980s. The general and the issue publics decisively affected the coalition-building processes within the party system, which in turn considerably narrowed the range of options available to policymakers. However, the change has nothing to do with public opinion as such. For example, the public had been as ambiguous about the deployment of nuclear weapons in West Germany in the 1950s as it was in the 1980s, although the policy impact was markedly different in the latter case. Rather, the difference in influence seems to reflect a change in the domestic structure that opened up German politics to societal pressures. Adenauer's "chancellor democracy" changed into a "party democracy" during the 1960s, thereby establishing a democratic corporatist policy network. On top of that, new social movements—from the student protests of the late 1960s to the environmentalists of the 1970s and the peace groups of the 1980s—affected the political culture and were crucial in opening the party system to societal influences.[73] In contrast to France the domestic structure of the Federal Republic was open enough to allow for such changes in the first place. And unlike the United States, the strength of the party system in the policy network allowed for the institutionalization of the societal consensus.

[73] For analyses of these changes, see Kendall Baker et al., *Germany Transformed: Political Culture and the New Politics* (Cambridge: Harvard University Press, 1981); Samuel Barnes et al., *Political Action: Mass Participation in Five Western Democracies* (Beverly Hills, Calif.: Sage, 1979); Karl-Werner Brand et al., *Aufbruch in eine andere Gesellschaft: Neue soziale Bewegungen in der Bundesrepublik Deutschland* (Opladen: Westdeutscher Verlag, 1983); and Richard Stöss, ed., *Parteien-Handbuch*, 2 vols. (Opladen: Westdeutscher Verlag, 1983–84).

CONSERVATIVE LEADERS VERSUS PUBLIC OPINION: JAPAN

Given the Japanese domestic structure and the stability of Japanese pu
lic opinion, the case is expected to resemble Germany. If one consid
the low-key approach of Japanese military security policy over the l
forty years, the self-imposed arms-control measures such as the nc
nuclear principles and the limits on defense spending, and the reluctar
to play a more global role in world politics, Japanese public opinion a
foreign policy seem to be more or less in line. Indeed, almost every an
ysis of Japanese defense policy points to the constraining forces of pub
opinion on the margins of maneuver of the policymakers.[74]

At first glance, Japanese foreign policy during the 1980s seems to cc
firm the analysis. Prime Minister Nakasone met with firm domestic c
position when he talked about making Japan an "unsinkable aircr
carrier"[75] and tried to move Japanese security policy toward a more pı
defense stance and to initiate a major military buildup program. Wh
all Japanese governments of the 1980s declared that they wanted to aba
don the principle of not exceeding 1 percent of the GNP for defense e
penditures, they were only able to do so from 1986 to 1989. The princi
was reinstated in December 1989. Finally, Japanese policy toward t
Soviet Union remained in line with the comparatively high degree
public threat perception (see data above). Japan was reluctant to react
Gorbachev's peace initiatives and to the changes in Soviet foreign polic
In contrast to the Federal Republic of Germany, Tokyo only grudging
accepted the INF treaty after having insisted that the Asian-based Sovi
SS-20s be included in the zero option.[76]

However, this description gives only half of the picture. The very fa
that Nakasone and his successors were able to change Japanese defen
posture as well as to implement a more globally oriented foreign poli
belies the notion that Japanese policy-making is dominated by trends
public opinion. Moreover, the 1 percent GNP limit on defense procur

[74] See, for example, J. W. M. Chapman, R. Drifte, and I. T. M. Gow, *Japan's Quest f*
Comprehensive Security (New York: St. Martin's Press, 1982); Scalapino (fn. 23); and Welfie
(fn. 24).

[75] In March 1983, 72% of Japanese were worried about Nakasone's defense policy; 48
disagreed with his remarks about the "unsinkable aircraft carrier." While public oppositi
to Nakasone's policies declined somewhat over the years, he was nevertheless unable
change the prevailing trends. For data, see Hastings and Hastings, eds., *Index to Internation*
Public Opinion, 1982–1983 (New York: Greenwood, 1984), 316. For analyses of Japane
foreign policies during the 1980s, see, e.g., Drifte (fn. 23); idem, *Japan's Rise to Internation*
Responsibilities (London: Athlone Press, 1990); and Malcolm McIntosh, *Japan Re-Arm*
(London: Frances Pinter, 1986).

[76] See details in Drifte (fn. 75, 1990), 48–58.

ment seems to be largely symbolic and primarily meant for domestic consumption. If one uses NATO criteria to measure military expenditures, Japan had the third highest military budget worldwide in 1988. Finally, Tokyo's tough-minded policy vis-à-vis Gorbachev's Soviet Union reflected public opinion on only a superficial level. As the above-quoted data suggest, most Japanese would have been perfectly willing to support a policy that tried to resolve the grievances with the USSR—particularly the dispute over the northern territories—by means of diplomacy.

The contradictory evidence is explainable: simply put, the LDP's conservative leaders did not share the general public's outlook on Japan's role in the world and moreover felt under constant U.S. pressure to build up the military. In essence, security policy seems to have remained the ne issue-area in Japanese politics in which the beliefs of the leaders of ie conservative coalition did not accord with public attitudes. As a result, Nakasone and his successors had to conduct their defense policy gainst the prevailing trends in public opinion; unable to change it and ven Japan's domestic structure, however, they had to acquiesce to it. hey thus muddled through while stretching public tolerance to its limits. If one accepts the notion that the Japanese governments indeed felt ılnerable to public opinion, the strange coincidence of low-key rhetoric xcept for Nakasone's early years) and considerably more active security olicies becomes explicable.

Thus, in the 1980s Japanese public opinion placed broad, but neverieless effective, constraints on the actions of policymakers. The impact the public was reinforced by the peculiarities of the Japanese domestic ructure, in particular, the emphasis on consensus building in the policy twork. Unlike France and more comparable to Germany's corporatist rocess, issue publics and societal actors were able to affect the coalitionilding processes within the LDP by forcing it to take account of the ewpoints of the opposition parties. As a long-term consequence of the 960 crisis, the conservative coalition tried to include the opposition pares in major decisions concerning foreign and security policy. The instiitionalization within the LDP of "policy tribes" corresponding to the irticular ministries and Diet committees also enhanced the consensusilding capacity within the LDP and reduced the impact of factionalm.[77]

[77] I thank Nobuo Okawara for pointing this out to me. See also Calder (fn. 42), 463. The ost recent domestic turmoil in Japan on how to deal with the crisis and the war in the rsian Gulf confirms the analysis that mass public opinion severely contrains the country's ling conservative coalition.

CONCLUSIONS

The study leads to three major conclusions. First, *mass public opini mattered* in each of the four countries, albeit to very different degre Policymakers in liberal democracies do not decide against an ov whelming public consensus. In most cases, mass public opinion set bro and unspecified limits to the foreign policy choices. In the U.S. and more recently—Germany it also defined the range of options availal for implementing policy goals. Additionally, general issue salience seer to be of minor importance, once there is a clear majority in favor certain policies. In the Japanese case the issue salience of foreign poli was comparatively low; decision makers nevertheless felt the pressure mass public opinion. In sum, public consensus may substitute for a la of issue salience, whereas highly mobilized societal actors may be able influence policies even in the absence of such consensus. Neither was t case in France.

However, there are discernible limits to the impact of the general pu lic on foreign and security policies. Rarely does general public opini directly affect policy decisions or the implementation of specific polici In sum, the bottom-up and the top-down models are too simplistic to the reality in the four countries.

Second, most of the previous research on the policy impact of publ opinion treats the domestic decision-making process as a black box an directly compares opinion polls with policy outcomes. This study su gests, however, that the *indirect effects of public opinion* are far more im portant. To the extent that the empirical material was available,[78] it coul be shown that the main role of the public in liberal democracies is influence the coalition-building processes among elite groups. In the U.S and the Federal Republic of Germany public opinion had a crucial im pact on elites as they reconfigured their coalitions; groups in line wit public preferences were able to carry the day. In Japan the public force the ruling conservative coalition to integrate the opposition into th decision-making process, at least to some extent.

Moreover, support by mass public opinion seems to be essential fo issue publics, public interest groups, and other societal actors to influenc policy decisions. This accounts for the partial success of the peace move ments in Germany and the U.S. and of the otherwise marginalized op position parties in Japan in affecting the debates in the policy network

[78] This proved to be a major limitation of this study, particularly in the French case. Th prevailing realist paradigm in international relations and the focus of most studies on elite do not encourage scholars to trace the impact of public opinion and societal actors on deci sions in detail.

For both the political elites and societal actors, mass public opinion proves to be a resource for strengthening one's position in the coalition-building process.

The third conclusion concerns the usefulness of the *domestic structure approach*. Domestic structures seem indeed to account for the differences between the four countries. The degree of mass public consensus on foreign policy correlates strongly with the degree of *societal fragmentation*, in particular, the ideological divisions between Left/liberals and Right/conservatives. Japan as the most homogeneous society of the four countries also enjoys the most stable public foreign policy consensus. By contrast, the divisions in French public opinion on security policy reflect the country's ideological and class cleavages. The U.S. and Germany are cases in between.

Furthermore, there is strong empirical evidence that domestic structures are the intervening variable between public opinion and foreign policy. Under given international conditions and despite relatively similar public attitudes across countries, variances in the interaction between the general public and elites in the foreign policy–making process can be explained by differences in domestic structures. The degree to which political institutions are centralized and the degree to which the state dominates the policy networks seem to be the determining factors. In both categories, France and the U.S. are located at opposite ends of the spectrum. The two countries also form the two cases in which public opinion seems to have the greatest (U.S.) and the least (France) impact on policy decisions. The Federal Republic of Germany changed over time from top-down to a more interactive or even bottom-up process, which in turn seems to reflect changes in the overall domestic structure. In the Japanese case the stable public consensus in favor of a nonmilitary security policy severely constrained a political elite that tried to pursue a more active foreign policy. Moreover, societal protest such as was experienced during the 1960 crisis left a discernible mark on the consensus-building process in the conservative coalition and, as a result, on Japanese foreign policy.

At first glance, the empirical findings appear to confirm the distinction between strong and weak states. The stronger the state institutions and the greater their control of the policy networks, the less access the public has to the policy-making process. However, while the strong versus weak state distinction seems to make sense regarding the *input* side of the political system, it is flawed if one considers the relationship between societal demands and *policy outcomes*. The strong French state and corporatist Japan both pursued comparatively low-key and noncommittal

policies with respect to the Western alliance system, arms control, a
the Soviet Union during the 1980s, but for very different reasons. Fre
foreign policy reflected not so much state strength as the fragility of
elite consensus and a general lack of policy legitimacy in the absence
societal input into the decision-making process. But Japanese fore
policy was constrained by the institutionalization of the domestic cons
sus—elite and public—in the policy network.

The comparison between the weak U.S. state and corporatist G
many is also revealing. Both countries pursued comparatively active a
highly visible foreign policies (Germany at least in the East-West co
text). But there are important differences that reflect the respective
mestic structures. The frequent volatility and unpredictability of Am
ican policies seems to result from constantly shifting coalitions
Washington's domestic process. In comparison, the Federal Republic
pursued a more stable foreign policy over the last twenty years. In co
trast to the U.S., but similar to Japan, the German policy network a
the political structure emphasize consensus building and the institutic
alization of coalitions in support of specific policies. As a result, the G
mans have less direct input into the policy-making process than t
Americans have, but their impact on policy outcomes seems to l
longer.

I am aware that this paper suffers from a number of limitations a
that further comparative research is indicated. First, while this arti
more or less covers the spectrum of domestic structures in liberal demo
racies, it deals neither with authoritarian regimes nor with one-pa
systems. Does the empirical evidence confirm the conventional wisdo
that public opinion does not count in those countries, or do systems th
are not democratic in the Western sense provide the masses with diffe
ent ways to influence policies? Second, the role of cultural factors to f
cilitate or to complicate the public's impact on foreign policy has to
examined systematically. The emphasis on consensus building as a soci
value in both Japan and Germany suggests, for example, that the dome
tic structure approach as presented here might have to be expanded. F
nally, this paper does not deal with the role of the media, though it
fair to assume that the media play a critical role both in the process
opinion formation and in conveying public attitudes to the policymaker

BRINGING CAPITAL BACK IN, OR SOCIAL DEMOCRACY RECONSIDERED

Employer Power, Cross-Class Alliances, and Centralization of Industrial Relations in Denmark and Sweden

By PETER SWENSON

SOCIAL democracy in power, according to the prevailing wisdom of over a decade of comparative study, needs unified, encompassing, and centralized labor unions. With labor's "relative power resources" so generated and mobilized in Scandinavia, Social Democrats could come to power in the 1930s and exact favorable terms from capital. According to this line of thought, after years of industrial strife of an intensity rarely experienced elsewhere in the world, the labor movement in Scandinavia gained impressive terrain with redistributive social policy envied by the Left elsewhere. Strikes practically vanished, rendered obsolete by legislation, a superior redistributive device.

This article aims to show that much of the literature on social democracy views power relations between classes more or less in this manner and errs in doing so. It critically examines what might be called the "balance of class power model"—most energetically and influentially developed by Walter Korpi[1]—which incorrectly depicts capital in the social democratic political economy as having been politically tamed by strong and militant labor confederations. It also advances another way of looking at the political regulation of class relations, drawing from the history of Danish and Swedish labor politics. This analysis suggests that centralized unions and the Left exercise institutionalized power in advanced

[1] See Walter Korpi and Michael Shalev, "Strikes, Industrial Relations and Class Conflict in Capitalist Societies," *British Journal of Sociology* 30 (June 1979), 164–87; idem, "Strikes, Power, and Politics in the Western Nations, 1900–1976," in Maurice Zeitlin, ed., *Political Power and Social Theory*, vol. 1 (Greenwich, Conn.: JAI Press, 1980); and Korpi, *The Democratic Class Struggle* (London: Routledge and Kegan Paul, 1983). See also distinct critiques of Korpi and related authors in Jonas Pontusson, "Behind and beyond Social Democracy in Sweden," *New Left Review* 143 (January–February 1984); and James Fulcher, "Labour Movement Theory versus Corporatism: Social Democracy in Sweden," *Sociology* 21 (May 1987), 231–52, esp. 237–42.

industrial societies not against or in any way at the expense of employe
In fact, the Danish and Swedish Left secured power in tacit alliance wi
dominant groups of employers. These employers indeed were not at
unwilling to construct the centralized institutions of industrial relatio
that are widely agreed to be constitutive features of labor-inclusive po
tics. This alliance perspective suggests there were winners and lose
within labor, as well as divisions among capitalists.[2]

The "cross-class alliance" model offered here provides one importar
if by no means the only, basis for explaining why Social Democratic ru
endured for so much of the interwar and postwar periods in Denmai
and Sweden. The political domination of Social Democrats, I argue, w
stabilized by the absence of intense opposition by capital to policies an
programs aggressively opposed by business elsewhere in the world. Th
quiescence was not a symptom of weakness or dependency. Instead,
was a product of the class-intersecting, cross-class alliance behind inst
tutions of centralized conflict resolution that routinely served mutual ir
terests of sectoral groupings that dominated employer and union confec
erations. In class-divisive, cross-class alliances these groups mutuall
reinforced each other's power to control intramural competitors and op
ponents. The argument disaggregates classes to focus on conflicts withi
as well as between capital and labor, and on interests shared across clas
lines in institutional systems regulating conflicts along both axes.[3]

With a focus on employers, intraclass conflict, and centralization, th
paper also sheds light on a hidden factor behind the cross-class (farmer
labor) parliamentary coalitions of the 1930s. Convention has it that stabl
governmental domination by the Left in Scandinavia was made possibl
by strong labor unions in combination with two other circumstances: (1
the advantages enjoyed by large, unified left-wing parties in parliamen-
tary contestation for power with small and divided centrist and right-
wing parties friendly to business, and (2) related to that, parliamentary
bargains with Agrarian parties, in which labor abandoned its opposition
to farm protectionism in exchange for farmers' acceptance of job pro-
grams in urban public works and housing development.[4] This is as far

[2] Note that the balance of power literature tends to see social democracy as a provider of positive-sum gains for both capital and labor. A coherent interpretation entails that capital's gains were imposed, not negotiated. Adam Przeworski's analyses of "class compromise" suggest a better alternative, as they emphasize the mutual advantages of institutional settlements rationally acted upon by both capital and labor. But he treats classes as unitary actors and therefore neglects intraclass conflict as a condition of cross-class settlements. See Przeworski, *Capitalism and Social Democracy* (Cambridge: Cambridge University Press, 1985).

[3] See also Peter Swenson, "Labor and the Limits of the Welfare State: Intra-Class Conflict and Cross-Class Alliances in Sweden and West Germany," *Comparative Politics* (forthcoming).

[4] See Francis Castles, *The Social Democratic Image of Society* (London: Routledge and Ke-

as the analysis of the cross-class alliances of the 1930s usually goes, which thus wrongly suggests that the coalitions excluded capital. In fact, for reasons to be discussed later, the alliance between sectors of labor and capital behind the centralization of industrial relations facilitated the formation of the more visible parliamentary farmer-worker alliances of the 1930s. In turn, the farmer-labor coalitions at the parliamentary level helped repair intraclass divisions that had been created by the industrial-level alliance. In other words, capitalists were essential participants—not outsiders—in this complex realignment of political forces.

This analysis of the coalitional foundations of social democracy opens with an account of the role of employers in coercing unions into institutions of centralized industrial relations in Denmark and Sweden. One of the chief instrumental motives of employers was to gain overarching ("peak-level") control of the intersectoral pay structure, mostly for the benefit of manufacturers of internationally traded metal products and to the disadvantage of high-pay workers in home-market sectors, especially in the building trades. Working in a sector sheltered from international competition, building tradesmen enjoyed advantages in wage bargaining not shared by workers in traded-goods sectors. This fact troubled manufacturers probably even more than it did building contractors. In a manner peculiar to Scandinavian employers, their associations went on the attack, using the sweeping multi-industry sympathy and offensive lockouts to promote peak-level control over, among other things, pay in the construction industry.

The paper highlights the militant role of employers in fostering union centralization, to show how the balance of class power model has systematically misconceived capital as passive or weak. An interpretation of employers' perceived interests and instrumental actions is integrated into a critical review of conventional explanations of such phenomena in Denmark and Sweden as (1) farmer-labor coalitions that brought the Left to power in Parliament and the bureaucracy from the 1930s onward and (2) the rapid decline in industrial conflict, also beginning in the 1930s. The paper concludes with a discussion of how labor's power and the cross-class parliamentary alliances of the 1930s should be reconceptualized in conformity with these revisions.

Explaining the Centralization of Labor: Denmark and Sweden

Organized employers wanted and aggressively promoted the centralization of industrial relations in Scandinavia. This idea is rarely entertained

gan Paul, 1978); and Gøsta Esping-Andersen, *Politics against Markets: The Social Democratic Road to Power* (Princeton: Princeton University Press, 1985).

in the literature, possibly because it is at odds with the prevailing and intuitively plausible equation of union unity and centralization with strength against capital. Comparative-historical case analysis of actual institution building in industrial relations, let alone the study of employers as architects and engineers in the joint project, are practically nonexistent.[5] On the "corporatism" question one typically finds ahistorical macrocomparative studies focusing on economic infrastructure, party politics, and the state—structural features of economics and politics that distinguish the centralizers from the others. Employers as organized agents are usually absent from the analysis and by implication inert.[6]

This criticism pertains primarily to the balance of class power literature, the specific features and problems of which are the focus of detailed analysis below. A different literature with varying Marxist angles on Social Democratic corporatism also ignores businessmen and institution building. The state here is supposed but never demonstrated to be the agency that imposes centralized industrial relations in order to harness labor unions to the capitalist wagon. With or without reference to instrumental state action, this literature seeks unsuccessfully to identify some systemic or functional logic behind the corporatist fusion of state and societal power. In any event the process is co-optative, operating at the expense of labor rather than of capital.[7]

[5] To explain variations in peak-level centralization, industrial relations expert John Windmuller examines and dismisses six factors; employers were not on the list. See Windmuller, "The Authority of National Trade Union Confederations: A Comparative Analysis," in David B. Lipsky, ed., *Union Power and Public Policy* (Ithaca, N.Y.: School of Industrial and Labor Relations, Cornell University, 1975). Keith Sisson, in *The Management of Collective Bargaining: An International Comparison* (Oxford: Basil Blackwell, 1987), focuses on employers to explain bargaining centralized at the industrial level but fails to explain peak-level centralization; and H. Clegg, *Trade Unionism under Collective Bargaining: A Theory Based on Comparisons of Six Countries* (Oxford: Basil Blackwell, 1976), offers an early but undeveloped statement on the importance of employers.

[6] The political economy literature ignores employer organization. Philippe Schmitter uses union organization as a proxy measure of corporatist organization in all other sectors; see Schmitter, "Interest Intermediation and Regime Governability in Contemporary Western Europe and North America," in Suzanne Berger, ed., *Organizing Interests in Western Europe: Pluralism, Corporatism, and the Transformation of Politics* (Cambridge: Cambridge University Press, 1981), 293. Peter Katzenstein treats the organization of labor and business independently, but he measures intersectoral concentration in peak trade organizations rather than centralization of authority in employer organizations, which are often separate. See Katzenstein, *Small States in World Markets: Industrial Policy in Europe* (Ithaca, N.Y.: Cornell University Press, 1985), 106. Trade, not employer associations, is the main focus of Schmitter and Wolfgang Streeck, eds., *Private Interest Government: Beyond Market and State* (London: Sage, 1985).

[7] According to Schmitter, Scandinavian style "societal corporatism" (which includes centralization) follows the gradual "decay of pluralism" and meets "certain basic imperatives . . . of capitalism to reproduce the conditions for its existence and continually to accumulate further resources." See Schmitter, "Still the Century of Corporatism?" in Schmitter and Gerhard Lehmbruch, eds., *Trends toward Corporatist Intermediation* (Beverly Hills, Calif.: Sage, 1979), 24. The state promotes corporatism in Leo Panitch, "Trade Unions and the

An employer-centered account of centralization and social democracy in Denmark and Sweden offers an alternative to both the balance of class power perspective on centralization and the more Marxist approaches. It focuses on the interests and instrumental actions of capitalists without endorsing the Marxist view that corporatism means co-optation of labor on behalf or at the behest of capital. In Sweden and Denmark certain factions within the Social Democratic labor movement desired centralization for their own material and power interests, but they lacked the muscle to impose it on reluctant unions without helpful and coercive pressure from employers. In alliance with labor interests, employers forced those reluctant unions to delegate authority upward to confederation leaders, primarily over the initiation and financing of strikes.

The Weapon: Lockout

Scandinavian employers used multisectoral "sympathy" and "offensive" lockouts to promote the centralization of union authority and industrial relations. When confronted individually by spontaneous, locally isolated militancy, organized firms would as a rule receive financial and other support during strikes against workers who could not manage for long without pay. In this way employers were partially immunized against uncoordinated action. When, however, a national industry or craft union stepped in with financial help in the classic whipsaw fashion, picking off one employer at a time, strike insurance proved of limited benefit (unless, as was often difficult, sufficiently numerous and skilled scabs could be procured). In response, Scandinavian employer organizations called upon their members to lock out many if not all remaining workers in the industry or craft. The employers generally received partial compensation from their organizations' conflict funds, while their idled workers in turn would demand lockout compensation from rapidly dwindling union funds. Duty-bound by their constitutions, the union confederations would then step in when called upon for aid, only to find that employers could counter with a multi-industry sympathy lockout of workers even in firms or industries not engaged in the disputes and operating peacefully (and often quite profitably) under valid union contracts. General lockouts of this nature struck confederation leaders at one of their weakest points—their finances. Until the 1940s Scandinavian employers periodically saw the need for such "bloodlettings," as they called it.

Capitalist State," *New Left Review* 125 (January–February 1981), 21–43; and Claus Offe, ed., "The Attribution of Public Status to Interest Groups," in *Disorganized Capitalism: Contemporary Transformations of Work and Politics* (Cambridge: MIT Press, 1985).

Sometimes employers simply skipped the intermediate steps and called sweeping multi-industry lockouts in response to strikes by no more than a handful of workers in a single craft or industry. In 1899, in a decisive move to force the Danish Labor Confederation (DSF) to restrain its member unions, the Danish Employers' Confederation (DAF) called a lockout first against workers in the entire carpentry trade and then in other unrelated industries, including engineering. Strikes by carpenters in seven provincial towns involving only a few hundred workers triggered the draconian response. Because the labor confederation was formally committed to defense of locked-out workers in the other sectors (with lockout support financed by ad hoc assessments of the entire membership), it was inevitably dragged into the conflict. After four months employers achieved a dramatic victory with the so-called September Agreement (Septemberforliget). The conflict practically bankrupted the unions, and because many previously unorganized employers rallied to join, the employers' confederation emerged financially better off than before.[8]

This peak-level agreement spelled out mutual recognition of the right of employers and workers to organize, along with employers' managerial prerogatives against worker control and the closed shop. Also, by requiring two weeks' advance notice of a strike or lockout vote, a three-quarters majority in the vote, and then another week before the conflict could begin, the agreement limited the number and kinds of strikes the labor confederation could be drawn into. Financially in ruins, the confederation quickly revised its constitution in 1900. The change empowered a three-quarters majority of its representative council to require a member union to conclude or prevent a work stoppage "when it has assumed, or threatens to assume, such dimensions that it endangers the peaceful continuance of work in the other trades."[9] Endangerment of other trades meant, presumably, provoking lockouts, the effect if not the purpose of which was to drag the unwilling confederation into expensive full-scale confrontations.

The years after 1900 saw only minor growth in the formal power of the DSF over its member unions' bargaining practices. It was to remain the least centralized of the Scandinavian labor confederations.[10] In practice, however, collective regulation of pay and working conditions be-

[8] This discussion on Denmark relies largely on Sophus Agerholm and Anders Vigen, *Arbejdsgiverforeningen gennem 25 Aar: 1896–1921* (The employers' confederation through twenty-five years) (Copenhagen: Dansk Arbejdsgiverforeningen, 1921), 3–142.

[9] Walter Galenson, *The Danish System of Industrial Relations: A Study in Industrial Peace* (New York: Russel and Russel, 1952), 59.

[10] Ibid., 2–3.

came highly if unstably centralized at the peak level because Danish employers, being the *most* unified and centralized, periodically used multiemployer and multisectoral lockouts to force formally autonomous unions to accept if not seek the central mediation and support of the DSF. The DAF indeed wanted further formal centralization of DSF authority, but it was forced to apply the disciplinary lockout repeatedly until the 1930s. Then Social Democratic governments began applying additional disciplinary force by regulating voting procedures in strike referenda and with compulsory and binding arbitration agreements.

In Sweden centralization of union power got off to a slower start but advanced further. From a position of impressive strength organized employers coerced the most recalcitrant unions to accept centralization. The key formative events occurred in the 1930s and bear interesting similarities to Danish events forty years earlier. Strikes in the high-wage building trades in 1933 were accompanied by threats of a sweeping multiindustry sympathy lockout by the Swedish Employers' Confederation (SAF). The leadership of LO (the Swedish Labor Confederation) quickly intervened, as this was no idle threat. Massive lockouts of similar nature had inflicted devastating financial and membership losses on LO (particularly after it was dragged into supporting a general strike in 1909).[11] Also pushing LO to gain control of its member unions, the Social Democratic government threatened statutory intervention. In 1938 at Saltsjöbaden LO and SAF signed a peak-level agreement that in many respects resembled the Danes' September Agreement of 1899. By signing at Saltsjöbaden, LO leaders agreed implicitly that, among other things, they should refuse support to striking unions (and for workers locked out in response) when the unions in question failed to abide by procedural formalities before taking action. Precipitous action and disregard for procedure was characteristic of highly successful militancy, especially in sectors like construction, whose unions opposed the agreement.[12]

In conjunction with negotiation of the 1938 agreement, and to secure the ability to enforce it, LO leaders formally responded to external (and internal) pressures to centralize control much as DSF leaders did after the September Agreement. Organizational reforms passed in 1941 gave LO the authority to veto all strikes involving more than 3 percent of an LO union's members, thereby strengthening their mandate to keep strike

[11] Bernt Schiller, *Storstrejken 1909: Förhistoria och orsaker* (The general strike of 1909: Background and causes) (Göteborg: Akademiförlaget, 1967).

[12] There was of course much more to the agreement. See Sven Anders Söderpalm, *Arbetsgivarna och Saltsjöbadspolitiken: En historisk studie i samarbetet på svensk arbetsmarknad* (Employers and the politics of Saltsjöbaden: A historical study of cooperation in the Swedish labor market) (Stockholm: SAF, 1980).

funds under lock and key. They also passed a resolution forcing member unions to incorporate LO's previously recommended standard bylaws (*normalstadgar*), which had been drawn up in 1933, the year of the construction strike. Most important, these banned binding membership referenda on strike and contract matters. (In the construction sector, the vote of the membership had frequently overridden more restrained leaders.) Construction and typographical unions spoke out against the 1941 reform but accepted it; refusal to do so would have meant expulsion and perhaps suicidal isolation in the face of the combined might of employers.[13]

THE MOTIVE: REDISTRIBUTION

Employers' desire and means to corral unions into centralized structures were unmistakable and similar in both Denmark and Sweden. Comparative analysis of the internal politics of the employers' confederations suggests also a shared motive. The fact that strikes in the construction trades precipitated the developments in both countries is already a clue. As early as 1907, when Danish, Swedish, and Norwegian employers met in Copenhagen to discuss Nordic employer cooperation, one of the few substantive resolutions they made regarding wage policies was to "keep earnings in the building trades roughly in line with earnings in other industries."[14] By other industries is meant above all those active in or at least vulnerable to international trade (producers of traded goods)—metalworking in particular. Not subject to international price competition, home-market, or sheltered, employers could more easily pass on the costs of high wages to consumers. Indeed, control of the intersectoral pay structure was a prime motive for encouraging unions to centralize authority across industries, since short of statutory control, centralized peak-level coordination was the only sure way of maintaining control without frequent resort to the lockout. For the lockout had its costs of course—financial ones for employers' organizations, but also in terms of internal solidarity between those who at any particular time might be eager to lock out and those who could make peace and profits (at higher wages) simultaneously.

High and rapidly rising pay in the sheltered building trades often caused difficulties for internationally vulnerable metals employers in se-

[13] Axel Hadenius, *Facklig organisationsutveckling: En studie av Landsorganisationen i Sverige* (Union organizational development: A study of the confederation of labor in Sweden) (Stockholm: Rabén & Sjögren, 1976), 57–68, 126–30.

[14] Anders Vigen, *De Nordiske Arbejdsgiverforeningers Samarbejde gennem 20 Aar* (The Nordic employer associations' cooperation over twenty years) (Copenhagen: Langkjær, 1927), 33–34.

curing the labor and cooperation of painters, electricians, plumbers, woodworkers, carpenters, and others needed for shipbuilding, electronics, and the manufacture of railroad cars and automobiles. Paying them competitive wages often upset internal pay differentials within those sectors and aroused further militancy within the metalworkers' union. Further, high wages in construction led to higher fixed capital costs, and militancy could disrupt the timely expansion of manufacturing capacity to meet fleeting market opportunities. Building wages translated into higher rents, which in turn generated greater pressure for wage increases outside of the building trades. Relatively low wages and expensive, scarce, and extremely cramped worker housing made it difficult to attract labor during expansion. Militancy spread across sectors to interrupt production and delivery in markets where buyers could easily be lost to international competitors poaching for new market shares. High wages could also mean the loss of domestic markets to imports.[15]

Though they were probably to benefit from it in the long run, employers in construction were not the driving force behind centralization of bargaining institutions. In intraorganizational politics they were the clear losers. In fact, employers in metals and engineering exposed to, though not particularly active in, heavy international competition were the most militant of all in pushing for centralization and the terms of the Danes' September Agreement. In the DAF's early years these employers fought an "open war" against Niels Andersen, the leader of the confederation (from the building industry); he was considered too moderate in his dealings with the unions and too willing to submit disputes to arbitration. But after the 1899 lockout and the September Agreement, and a revival of tensions and near repeat of the lockout in 1907, relations between employers in the metals and construction industries were calmed by the DAF's escalating militancy. In 1907 employers in the metals industry (accused by others of having "the lockout disease") achieved better representation in DAF, and by 1911, Andersen's most vocal critic, S. C. Hauberg (from the metals sector), had taken over leadership of the powerful confederation.[16]

Similar yet more drawn-out tensions between employers in home-market and internationally exposed industries were present in Sweden. Tensions finally eased in the 1930s, when SAF undertook to clear the con-

[15] See Karl-Olof Faxén, "Några kommentarer till SAFs lönepolitiska uttalanden under 1920-talet," in Eskil Wadensjö, Åke Dahlberg, and Bertil Holmlund, eds., *Vingarnas trygghet: Arbetsmarknad, ekonomi och politik* (The safety of wings: Labor markets, economics, and politics) (Lund: Dialogos, 1989).

[16] Agerholm and Vigen (fn. 8), 26–28, 42, 153, 159–61, 173–90, 283.

struction industry of a jungle of practices that facilitated successful guer rilla warfare by building tradesmen. By that decade export-oriented em ployers had established considerable power in SAF, though not withou losing a faction of building contractors that bolted from SAF about 1933 The process began in 1917, when independently organized engineerin employers joined SAF, trading in their autonomy for substantial contro within the confederation. Not surprisingly, the following year the SA leadership under Hjalmar von Sydow dropped its old misgivings abou absorbing construction employers; it turned to the more encompassin Danish and Norwegian employers' confederations for a model of how t control construction workers.[17] SAF's hopes were dashed when employer in construction settled in 1924 for high increases and, even worse, a sys tem of piece-rate setting that would have severe wage-inflationary con sequences in the 1930s.[18] Despite the depression, wages and militanc continued to rise, uncontrolled by unions and employers in the buildin industry.[19]

During the depression employers in traded goods were forced to re duce prices to maintain production, and unions like the metalworkers accommodated them by accepting wage reductions. Their ability to do so was limited by the fact that the metalworkers' militancy could not be contained long if pay differentials across sectors were not reduced and kept under control. The leadership of the metalworkers' union had proved its reliability, having recognized the costs of militancy and strikes in terms of lost jobs and pay resulting from the loss of unrecoverable market shares in intense international competition. It could not maintain long-term restraint, however, if wages in home-market sectors were not somehow indirectly subjected to the same discipline.[20]

Employers in SAF threatened a sweeping economy-wide lockout in response to strikes in the building trades begun in 1933. The Social Democratic government also threatened intervention. Unions representing members paying high rents and suffering from job and housing shortages naturally had mixed feelings about LO fulfilling its obligations to support construction workers.[21] As a consequence of pressure from So-

[17] Centrala Arbetsgifvareförbundet, *En Minnesskrift, 1903–1918* (Memorial publication, 1903–1918) (Stockholm: Centrala Arbetsgifvareförbundet, 1919), 35–36.

[18] Carl Hallendorff, *Svenska Arbetsgifvareföreningen* (The Swedish employers' confederation) (Stockholm: SAF, 1927), 158–61.

[19] For reasons, see especially Hartmut Apitzsch, "Byggnadsbranschen: Produktionsförhållanden och organisationsstruktur" (Construction: Production relations and organizational structure), *Arkiv för studier i arbetarrörelsens historia* 2 (1972).

[20] For a lengthier discussion and references, see Swenson, *Fair Shares: Unions, Pay, and Politics in Sweden and West Germany* (Ithaca, N.Y.: Cornell University Press, 1989), 42–53.

[21] Bo Carlson, *Trade Unions in Sweden* (Stockholm: Tiden, 1969), 37.

cial Democrats, Agrarians, and SAF, LO finally prevailed on the construction unions to call off the strikes. The confederation stepped in again with even less hesitation in 1937 with a "brutal public rebuke" of two of its own construction unions that had called strikes.[22] For employers interested in maintaining international competitiveness by keeping sheltered wage-setting under control, this was icing on the cake from a labor movement that was already offering itself as an ally against militant syndicalists and communists. Moreover, unions like the metalworkers' were proving increasingly pliant about rationalization and automation. With enemies like these, who needed friends?

Small Cold States and Cross-Class Alliances

For employers in small economies dependent on or vulnerable to international trade, problems with regulating pay across home-market and traded-goods sectors were probably more acute than they were elsewhere. In large countries employers in traded goods had larger home markets and were therefore better insulated from international discipline. They would have been better able to raise prices along with wages to keep up with the construction industry. Their ability to withstand strikes may also have been greater when they resisted following construction wages; that is, using their *national* employer organizations, trade associations, and perhaps connections to centralized banks, they could coordinate punitive action against domestic competitors who poached on market shares made available by strikes. They could also coordinate domestic efforts to choke off supplies to employers making unacceptable agreements with unions; and they could more easily arrange the sharing of workers or facilities from far and wide to fill orders.[23] Internationally competing employers in small countries lacked large home markets, had no punitive devices against foreign competitors who took advantage of them when they were downed by strikes, could not enlist foreign businessmen to boycott employers making unacceptable deals with unions, and for reasons of scale (and no doubt because there were fewer regional and ethnic obstacles to solidarity) had a smaller pool of workers available as scabs.[24]

[22] Klas Åmark, *Facklig makt och fackligt medlemskap: De svenska fackförbundens medlemsutveckling 1890–1940* (Union power and union membership: Swedish unions' membership development) (Stockholm: Arkiv, 1986), 153.

[23] On the use of some of these practices in the U.S., see Clarence Bonnett, *Employers' Associations in the United States: A Study of Typical Associations* (New York: Macmillan, 1922), 45, 74–75, 109, 116, 140.

[24] Adolf Weber mentions German employers' problems in coordinating solidarity with foreign manufacturers to limit poaching, help fill orders, blacklist strikers who cross borders, procure scabs, and generate financial help; see Weber, *Der Kampf zwischen Kapital und Ar-*

For internationally competing employers in northern climates, the problems of intersectoral pay deviations, and therefore need for centralized control, were especially acute. Concerned about the problem caused by high differentials, engineering employers in Sweden compiled crude but nevertheless suggestive evidence showing that in 1930 differentials between metals and construction sectors in Scandinavia were substantially greater than they were elsewhere in Europe. One reason for the high differentials probably lay in the long Scandinavian winters, where low temperatures and short daylight hours make construction slow and costly. (See Table 1 for engineering employers' wage data and temperature conditions.) Mixing, placing, and curing concrete and brick or stone mortar at freezing temperatures and after an early nightfall required extra heating, lighting, and shelter.[25] To compensate for winters with little or no pay when construction ceased, workers in the trade demanded high wages and had the bargaining power to wrest them from contractors pressured by anxious speculators and manufacturers eager to finish their apartment buildings and industrial plants during the short building season.[26]

TABLE 1
HOURLY EARNINGS FOR MASONS AS A PERCENTAGE OF PAY FOR TURNERS AND FILERS IN SIX EUROPEAN CITIES (1930)

	Pay Differential	*Temperature (F)*[a] *(December–February)*
Stockholm	216	23.7
Copenhagen	152	29.7
Amsterdam	109	34
Warsaw	98	23
Rome	95	39.7
Belgrade	63	28

SOURCES: George Styrman, *Verkstadsföreningen 1896–1945* (The engineering association, 1896–1945) (Stockholm: Norstedt, 1946), 228; H. McKinley and Linda L. Liston, *The Weather Handbook* (Atlanta, Ga.: Conway Research, 1974).

[a] Average daily minimum, measured over 25–30 years.

beit: Gewerkschaften und Arbeitgeberverbände in Deutschland (The struggle between capital and labor: Unions and employer associations in Germany) (Tübingen: J. C. B. Mohr, 1921), 195–96.

[25] According to one estimate, there could be a fivefold increase in labor hours per cubic yard for mixing, placing, and curing concrete over what was required in ordinary weather, due to extra waiting time and use of heaters and protective enclosures. See H. E. Pulver, *Construction Estimates and Costs* (New York: McGraw-Hill, 1947), 137; and also Ralph P. Stoddard, *Brick Structures: How to Build Them* (New York: McGraw-Hill, 1946), 47–50.

[26] The relationship between temperature and high differentials in Table 1 does not show

Competing internationally from small northern countries, Danish and Swedish employers thus pushed hard for centralized, intersectoral control, especially over construction wages. The Swedish experience of the 1930s shows that employers had allies behind centralization—those with related interests, such as farmers. Most important to emphasize here are the interests of *unions* outside of the construction sector, in particular, metalworkers. Their leaders supported centralization because of internal membership discontent, when, in the face of militant employers and unemployment, they negotiated wage reductions while construction wages raced ahead. Unions pressed by their membership to level wages within as well as across sectors could not at the same time keep the wages of skilled workers in line with wages in construction without compensatory and divisive reductions of wages for unskilled workers.[27]

In Denmark another, somewhat different cross-class alliance underlay early centralization. Employers in engineering also wanted to contain construction wages. Again the union of skilled metalworkers was a key supporter of centralization. In this case, however, the most vocal opponents of centralization in the decades after the 1899 confrontation were probably not the construction unions representing skilled craftsmen; they were rather unskilled laborers—including both metalworkers and construction laborers—who unlike their Swedish counterparts were organized separately in their own general union. While the Swedish Metalworkers' Union representing both the skilled and the unskilled pushed successfully for interoccupational leveling, skilled Danish metalworkers supported employers' efforts to maintain differentials. Hence, playing a dominant role in the DSF and bitterly opposed by the laborers' union, they supported centralization and implicitly backed the employers' confederation against the militant but weak union of unskilled workers.[28]

In both countries, then, there was a class-splitting, cross-class alliance between employer and union groups in support of centralization. Employers in both countries pushed for centralization, using lockouts and threats of lockouts to force resistant but isolated and therefore economically vulnerable unions into the arrangement. In Sweden, as the following discussion shows, the Social Democratic Party and the government joined the alliance against construction workers. In Denmark, Social

up for Poland and Yugoslavia, where—to speculate—perhaps less mass production in metalworking, less urban-industrial construction, more repressive treatment of unions, and alternative livelihoods for construction workers in winter months affected relative wages differently.

[27] For more on the unions' distributional dilemmas (the "trilemma") and their related attitudes toward bargaining structures, see Swenson (fn. 20).

[28] See Galenson (fn. 9).

Democrats intervened more against the unskilled general laborers union, which was bitterly critical of the party, allied as it was with skilled metalworkers who favored centralization. Construction tradesmen do not figure in the vocal opposition, since employers *alone* had already brought relative pay for skilled building tradesmen under control and since further savings on building wage costs could be squeezed separately out of unskilled laborers.

Competing Views on Centralization and Implications for Understanding Social Democracy

The comparative literature of the last decade reveals considerable consensus about the institutional and coalitional foundations of the social democratic political economy. According to this consensus, a changing balance of power or distribution of power resources between capital and labor as undifferentiated blocs explains changes in institutionalized class relations—to the overall benefit of one and at the expense of the other.

This now-conventional analysis of institutionalized "labor inclusion" can be broken down into two somewhat contradictory historical renditions, which are criticized below. According to the first, politically strong and centralized union confederations emerged after and because Social Democrats achieved power in alliances with farmers that excluded industrial capital. The second view more or less reverses the sequence and holds that centralized unions gave labor the power resources to achieve and maintain parliamentary power at the expense of capital—or at least against the resistance of capital.

In either case, the Left parlayed power resources from one arena into advances in the other. Both versions support the widely held view, also challenged below, that the precipitous decline in industrial conflict in the 1930s can be explained by the achievement of parliamentary power backed by organizational centralization; that is, labor could achieve by legislative means what once required strikes. Drawing on the foregoing case-historical analysis, the following discussions deal with these contentions, showing that each is unfounded and that the causal connections between the related phenomena are better represented by the cross-class alliance model.

Red-Green Alliances, Capital Exclusion, and Corporatism

According to one version of the power balance model, once Swedish Social Democrats came to power with help from the Agrarian Party in 1932, employers were forced to back off and seek cooperative re-

lations with labor. Simply stated, according to John Stephens, "Swedish 'corporatism' was the result of the labour movement's strength."[29] Gregory Luebbert states the argument most explicitly in his comparative-historical explanation of different political realignments and regime changes in interwar Europe. In his view, the majoritarian alliances of Social Democrats and small farmers in Scandinavia and Czechoslovakia "created a stable power base" on which "they could build the corporatist institutions that consolidated their hold on the state and the working class." And the resulting centralized, encompassing unions "offset the otherwise inherent advantages of employers in collective bargaining and in the polity." Luebbert then argues that "from this position of dominance, they could, in turn, induce the collaboration of the business community."[30] He suggests that by excluding liberals, labor alliances with farmers excluded corporate capital as one of the "social foundations" of interwar social democracy.[31]

On the surface the Swedish case would seem to support the idea that the Social Democratic victory might have scared employers into centralized arrangements with unions, thus perhaps bestowing upon the labor confederation power never before enjoyed. According to a widely held view, employers in the 1930s dropped demands for state control of labor because conservatives could no longer be counted on to control the legislative process. Friendly compromise now appeared superior to legislation for achieving labor peace.

Certain facts cast doubt on this account. Even *before* the Social Democrats came to power in 1932, dominant employers in SAF feared state intervention—beyond what they happily accomplished in 1928 by installing a labor court with the power to award damages for violations of negotiated procedural and peace agreements. After 1928 ambivalent employers desired additional legal control primarily over strike and boycott tactics of syndicalists and communists. Responding to pressure from the Right, Social Democrats obliged by proposing just such legislation, which they withdrew in 1936 when their Agrarian allies lost interest.

[29] John Stephens, *The Transition from Capitalism to Socialism* (London: Macmillan, 1979), 135. To be fair, Stephens's analysis is eclectic: he correctly attributes earlier steps in centralization (pp. 42–45) to lockouts but also to "industrial infrastructure," which is wrong (see pp. 531–35 below).

[30] By saying capital was "induced to collaborate" Luebbert indicates incorrectly that capital was dominated, reluctant, and disadvantaged—if not fully "excluded." See Luebbert, "Social Foundations of Political Order in Interwar Europe," *World Politics* 39 (July 1987), 449–78, at 463–64.

[31] In the same tradition Ronald Rogowski also depicts the social democratic coalition, at least in Sweden, as capital exclusive; see Rogowski, *Commerce and Coalitions* (Princeton: Princeton University Press, 1990).

The Social Democrats, who had made no move to prevent the massive lockout threatened in 1933, included no measures in labor legislation to disarm employers—unlike New Deal Democrats in the U.S.! For these reasons, it hardly makes sense to depict employers' continued moves toward Saltsjöbaden after 1936 as any kind of political retreat extorted by threats of legislation.[32]

In the case of Denmark the partisan realignment of the 1930s cannot possibly account for strengthening centralized union power, which occurred much earlier under employer pressure. On the surface, however, at least one aspect of the Danish case might seem to confirm the picture drawn by Luebbert and others of social democracy as a coalition excluding or even hostile to capital. With the Kanslergade Agreement of 1933, the Agrarian Party gave the Danish Social Democrats the majority they needed to block the offensive lockout with which employers planned to impose a 20 percent wage reduction across the economy. As in Sweden, so in Denmark farmers gave Social Democrats the support necessary to stimulate the economy with a job creation program. In exchange, farmers demanded agricultural subsidies and because of the importance of agricultural exports, a 10 percent devaluation of the Danish crown.

Walter Galenson points out that the agreement "was received with bitterness by the employers, who denounced the Agrarian Party for delivering a 'brutal blow' to business interests under pressure from its semi-fascist extremist wing."[33] There are reasons to think however that this protest was in part symbolic (possibly to placate home-market employers). The Kanslergade Agreement did not lay waste the entirety of "business interests," since for employers vulnerable to foreign competition, a 10 percent devaluation could partially compensate for the disappointment of not achieving the 20 percent wage cut. This would have been true especially for capital-intensive, high-value-added industries using relatively few imported inputs.

At the same time, Danish manufacturers were favored by exchange controls, which gave them privileged access to foreign currency (especially for the purchase of British coal). Due to this pattern of advantages, in combination with quantitative import restrictions introduced in early 1932 and a strategy of import substitution via exchange controls, Danish manufacturing output increased so rapidly that the decade of the Great

[32] See, e.g., Steff Andersson, *Mellan Åkarp och Saltsjöbaden 1923–1928* (Between Åkarp and Saltsjöbaden) (Stockholm: Almqvist and Wiksell, 1990).

[33] Galenson (fn. 9), 131–32. Katzenstein (fn. 6) repeats the consensus in saying that the agreement "was directed against the Conservatives and their supporters in the business community" (p. 142).

Depression has been described as the "second wave of Danish industrialization."[34] Students of Latin American import-substitution strategies of the same period will see striking similarities to the Danish politico-economic realignment.

Manufacturers of traded goods in the Danish employers' confederation also gained from the tradition of government wage intervention initiated by the red-green coalition at Kanslergade. Repeated state intervention guaranteed intersectoral and interoccupational pay regulation substantially in harmony with what employers had already tried to impose with centralized control via the lockout. Soon after Kanslergade the coalition began using the new mechanism of compulsory arbitration against *unions* rather than employers, most notably against laborers but also, in 1939, against typographical workers. It is telling that this high-pay, home-market industry was the same industry whose unions in Sweden joined construction unions in resisting the institutional changes of the 1930s.[35]

As regards Denmark, therefore, the power of centralized unions was not generated by the red-green alliance and then applied to "induce the collaboration of capital," as Luebbert suggests. Compulsory arbitration following Kanslergade probably did more to freeze the existing crafts-based division of Danish (as compared with Swedish) unions than to further consolidate and centralize them.[36] Industry's collaboration with the Social Democrats was induced neither by the incentives nor by the threats that union centralization could deliver but rather by favorable legislation and administrative practice conforming to the interests of vulnerable manufacturers. Capital—at least a segment of it—was clearly included as a "social foundation" of Denmark's interwar regime.

[34] Hans Christian Johansen, *The Danish Economy in the Twentieth Century* (London: Croom Helm, 1987), 53–57. See also Erik Rasmussen, *Velfærdstaten på Vej 1913–1939* (The coming of the welfare state 1913–1939) (Copenhagen: Politikens Forlag, 1965), 416–28; and Harry Haue, Jørgen Olsen, Jørn Aarup-Kristensen, *Det ny Danmark, 1890–1978: Udviklingslinier og tendens* (The new Denmark, 1890–1978: Development patterns and tendencies) (Copenhagen: Munksgaard, 1980), 165–74.

[35] Galenson (fn. 9), 131.

[36] On the Czechoslovakian farmer-labor coalition, Luebbert cites Harry Klepetar, who in reference to the effect on unions only mentions compulsory arbitration and legislative extension of arbitrated settlements to unorganized firms. See Klepetar, *Seit 1918: Eine Geschichte der Tschechoslowakischen Republik* (Since 1918: A history of the Czechoslovakian republic) (Moravska Ostrava: Julius Kittls, 1937), 351, 357. Such intervention elsewhere in Europe weakened unions by making them dependents of the state and increasing employer antagonism. On interwar and postwar Germany, see Gerald Feldman and Irmgard Steinisch, *Industrie und Gewerkschaften 1918–1924: Die überforderte Zentralarbeitsgemeinschaft* (Industry and unions, 1918–1924) (Stuttgart: Deutsche Verlags-Anstalt, 1985); and Karl Anders, *Stein für Stein* [Stone by stone] (Hannover: Verlag für Literatur und Zeitgeschehen, 1969), 264–66. On interwar France, see Walter Kendall, *The Labour Movement in Europe* (London: Allen Lane, 1975), 74.

The encompassing character of the cross-class alliances in Scandinavia can also be seen in the way the distributional aims of centralization shared by union and employer groups alike also facilitated the parliamentary farmer-labor coalitions in the two countries. There was more to the Swedish and Danish farmers' willingness to join Social Democrats in these cross-class coalitions than the newfound sympathy they received for agrarian protection. Farmers also stood to benefit from the wage restraint that the centralized and disciplined unions exercised with the help of fraternal pressure from Social Democrats. Even small farmers typical of Scandinavia relied on laborers, sometimes only a few and only seasonally, but no less critically during peak harvest, haymaking, and planting and slaughter times. High pay in construction and resulting pay increases elsewhere tended to make farm workers increasingly expensive and prone to join the urban labor force at the worst possible moments.

For the leadership of the Agrarian Party in Sweden the most difficult thing to accept about the Social Democratic crisis plan in 1933 was its insistence that public housing projects pay the standard union rates, which threatened to jump too far ahead of farm wages. With their hopes staked on publicly financed housing construction as the solution to unemployment, LO and Social Democrats applied direct pressure to construction unions to accept defeat after their ten-month-long strike. Only then would the Agrarian Party agree to release government funds for the housing program.[37] Another intervention in 1937 and the Saltsjöbaden Agreement in 1938 reconfirmed for the Agrarians the reliability of LO and the Social Democrats.[38]

Danish dairy and meat farmers, who were very active in international trade, were also brought into the cross-class alliance of the 1930s on the same terms that appealed to manufacturers of traded goods—devaluation and control of the interindustry and interoccupational pay structure. The Agrarians' trust in the Social Democrats as coalition partners could only have been enhanced by the fact that they supported the centralization following the September Agreement of 1899 and the taming of

[37] See Olle Nyman, *Svensk parlamentarism 1932–1936: Från minoritetsparlamentarism till majoritetskoalition* (Swedish parliamentary government, 1932–1936: From minority government to majority coalition) (Uppsala: Almqvist and Wiksell, 1947), esp. 108–12, 136–37, 158, 528–32.

[38] Peter Gourevitch collapses the 1933 crisis agreement with the Agrarians and the 1938 agreement and incorrectly characterizes the Saltsjöbaden negotiations in 1938 as a time when farmers and government met with labor and employers. He misstates the nature of the agreement between labor and employers by asserting that labor gave up demands for socialization and agreed not to strike and that business "accepted high wages." In fact, business convinced the unions to accept wage *restraint* but not to renounce the use of strikes. See Gourevitch, *Politics in Hard Times: Comparative Responses to International Economic Crises* (Ithaca, N.Y.: Cornell University Press, 1986), 26, 141, 152, 179, 231.

workers in sheltered, home-market sectors like construction; this applied particularly to unskilled laborers, who easily moved from farmstead to building site.[39] Relative pay for Danish construction workers was well under control in the 1930s, while in Sweden the problem of severe volatility in relative wages was only solved that decade.[40] (See Figures 1 and 2.)

In sum, because employers in traded goods desired further centralization to regulate the intersectoral pay structure, it can only be concluded that capital was integrated rather than excluded from the centralizing cross-class coalitions of the 1930s. Social Democrats, through actual (Denmark) or threatened (Sweden) intervention, accomplished what employer power had failed to do: impose a more stable centralization to the satisfaction of farmers and manufacturers alike. This interpretation of events in the 1930s better fits the facts than does the argument that centralization was a consequence of growing left-wing political power, aided by Agrarian support—to the exclusion and at the expense of capital.

Late Industrialization and Industrial Structure

While some adherents of the balance of class power argument look mistakenly for causes of centralized industrial relations in the new political power of the Left in the 1930s, some seek the answer in the peculiarities of the industrial structure of small, late-industrializing open economies. If this latter brand of analysis were correct, it could show how unions acquired their centralized power independently of employers' distributional interests, extraordinary power, and instrumental actions—in order then to be able to make better sense of the idea that centralization could be used as a source of power *against* capital.

The most significant influence on this point of view has been the work of Geoffrey Ingham.[41] In Sweden and Norway, which best sup-

[39] Luebbert makes a similar and intriguing argument in no way inconsistent with mine that small landowning farmers in Scandinavia were less reluctant than were such farmers elsewhere in Europe to join with Social Democrats because the latter had failed to make socialists out of many farm laborers and therefore to undermine class domination in the countryside.

[40] Note that for Sweden much of the 1932 drop in construction wages is attributable to an improvement in the sample of employers reporting wages; the real drop came afterward. Calculations for Denmark are from average hourly earnings for skilled carpenters and metalworkers in Copenhagen, and average daily summer wages are for temporary male agricultural day laborers receiving board. For Sweden, average daily wages are calculated for all industrial workers, skilled and unskilled, and for male temporary day laborers paying their own board (summer only). See also Galenson (fn. 9), 176–77.

[41] Ingham, *Strikes and Industrial Conflict: Britain and Scandinavia* (London: Macmillan, 1974). An influential body of literature on comparative political economy has drawn on Ingham's thesis to help explain the economic and political peculiarities of small, open, export-

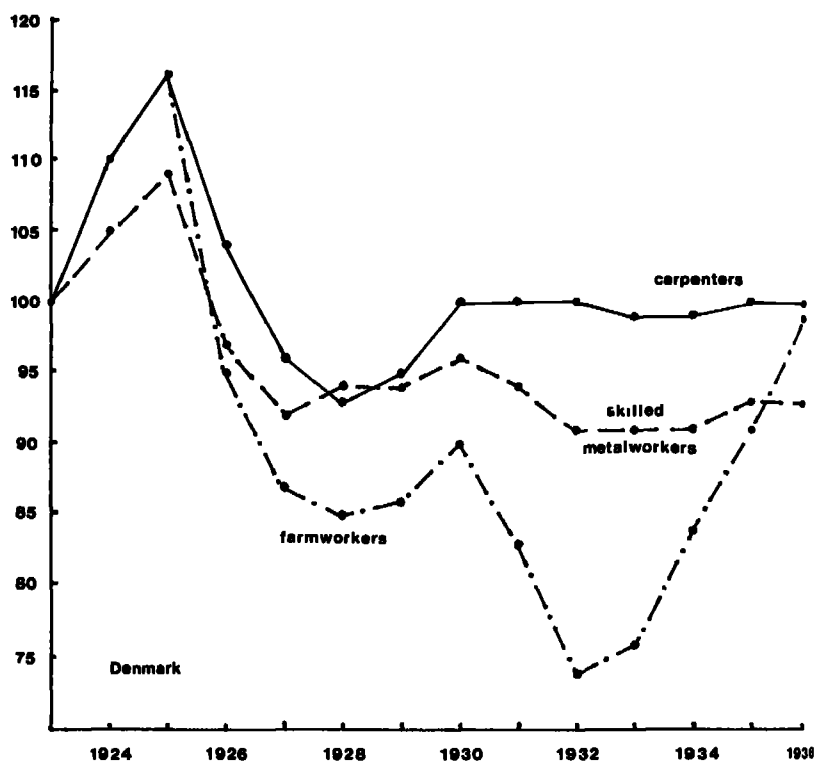



FIGURE I
DENMARK: NOMINAL WAGE TRENDS IN THREE SECTORS (1923 = 100)

SOURCE: Det Statistiske Departement, *Statistisk Aarbog for Danmark*, various years.

port Ingham's case, late industrialization and orientation toward large export markets produced highly concentrated industries based on mass production rather than on crafts technology. Concentration of ownership and control facilitated the efforts of large manufacturers to organize effectively, especially since they carved out specialized export niches and thereby avoided competitive rivalries. Their mass-production technology exploited economies of scale for competition in large export markets and created a relatively homogeneous working class undivided by crafts par-

oriented economies. See, e.g., David Cameron, "The Expansion of the Public Economy: A Comparative Analysis," *American Political Science Review* 72 (December 1978), 1243–61, at 1256–57; and Stephens (fn. 29), 42–44, 127, 141. In his analysis of the strong corporatist tendencies of small states, Peter Katzenstein does not dwell on the question of how or why employers and unions centralized and relies instead on Cameron and Stephens for their arguments. See Katzenstein (fn. 6), 91, 104.

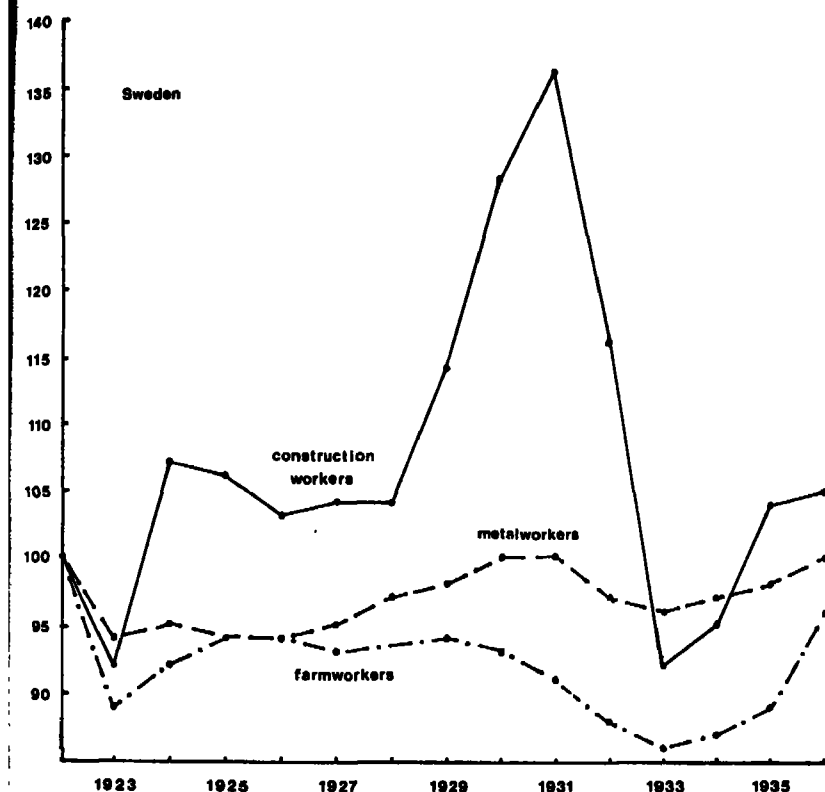



FIGURE 2
SWEDEN: NOMINAL WAGE TRENDS IN THREE SECTORS (1922 = 100)

SOURCE: Statistiska Centralbyrån, *Statistisk årsbok för Sverge*, various years.

ticularism. Centralized organizations of workers followed more or less naturally from homogeneity.[42] In Britain, by contrast, early industrialization and crafts-based small-scale enterprises left both employers and workers too numerous, diverse, and competitive to form strong centralized organizations.

The circumstances surrounding the formation of the Danish Employers' Confederation (DAF) in 1898 show the Ingham thesis to be entirely unsound. The *first* of the strong Scandinavian employer confederations to form, its actions led quickly to the early centralization of power in the Danish Labor Confederation (DSF). As founders of DAF, employers in the

[42] Ingham (fn. 41), esp. 35-36.

highly fragmented home-market industries dominated the confederation, especially masters' associations in the various building trades. Even the most militant advocates of greater centralization (the relatively small employers in metalworking) manufactured goods mostly for the home market. That is, employers centralized early and sought centralized industrial relations in the very country and sectors where Ingham's analysis would say it is least likely to be found—in easy-entry, small-scale, crafts-based, and home-market sectors. This observation accords with evidence from the industrial relations literature that centralized multiemployer bargaining is most common where there is intense competition among small firms in product markets where ratios of labor to capital are high. Aside from the obvious need to counteract unions' whipsaw tactics, a compelling motivation for such employers to seek or accept centralized pay regulation is to "take wages out of competition"—to prevent the scamping of labor from competitors by overbidding them on the labor market and also to inhibit price warfare and to set up an entry barrier to low-wage/low-price competition.[43] Competition, fragmentation, and even old technology associated with early small-scale industry helped promote employer solidarity and centralization rather than, as Ingham suggests, inhibit it.

For this reason, no doubt, Ingham strains to place Denmark in the company of Sweden and Norway. In the case of Denmark he focuses on lockouts and employers' coercive actions to impose centralization, whereas he neglects the same in Sweden and Norway. Ironically, in focusing on employer solidarity in the country where its infrastructural conditions should have been the least conducive to unity, Ingham departs radically from his main argument. Instead of relying on a force deriving directly from his infrastructural conditions (as he suggests he does), he identifies an autonomous explanatory force as decisive without locating either its infrastructural or motivational origins.

In a similar critique of Ingham's argument, Peter Jackson and Keith Sisson turn, as do Stephens and Luebbert, to politics as the cause of centralization. In their argument, which differs from that of Stephens and Luebbert, political division on the Right early in the century (rather than Social Democratic strength alone in the 1930s) forced employers to rely on their own centralized devices rather than on the state to control labor.

[43] See Frank C. Pierson, "Cooperation among Managements in Collective Bargaining," *Labor Law Journal* 11 (July 1960), 621–28; Jules Backman, *Multi-Employer Bargaining* (New York: New York University Institute of Labor Relations and Social Security, 1951), 15–18; and Lloyd Ulman, "Connective Bargaining and Competitive Bargaining," *Scottish Journal of Political Economy* 21 (June 1974).

Though plausible and historically better informed, their argument overestimates the importance of the 1902 general strike for worker suffrage in spurring centralized employer organization in Sweden. More important, it fails to identify decisive economic motives analyzed here that were present even without the political mobilization of labor and political division on the right.[44]

THE CONSEQUENCES OF CENTRALIZATION

One variant of the balance of class power interpretation has tended to draw on Ingham's explanation of centralization in order to use centralization to explain the relative political success of left-wing parties in electoral and parliamentary politics. David Cameron's work exemplifies best this integrative understanding of the causes and consequences of centralization.[45] His version, which differs distinctly from the work of Stephens and Luebbert, sees Social Democratic power as a consequence, and not a cause, of centralization. Although his argument about the consequences of centralization lacks secure footing in an explanation of centralization itself, it alone is plausible and influential enough to merit further analysis.

DO STRONG UNIONS EXPLAIN SOCIAL DEMOCRATIC POWER?

In probably the best of a number of cross-national statistical analyses of social democracy, Cameron finds that one factor—centralized, encompassing, and densely organized union confederations—seems to explain much of the variation in left-wing political control of government. He proposes a causal mechanism to connect the two: the "organizational resources" of unions or "labor's strength" relative to capital—an argument he borrows from Walter Korpi.[46]

[44] Jackson and Sisson, "Employers' Confederations in Sweden and the U.K. and the Significance of Industrial Infrastructure," *British Journal of Industrial Relations* 14 (November 1976). Walter Korpi challenges Ingham's data, arguing that if anything British industry is more concentrated in terms of the largest firms' share of total industry employment; see Korpi, *The Working Class in Welfare Capitalism: Work, Unions and Politics in Sweden* (London: Routledge and Kegan Paul, 1978), 394. Curiously, like Ingham, even Korpi relies on data from the 1950s and 1960s; and neither speculates that these data might reflect *consequences*, and not causes, of earlier bargaining structures. In fact, centralized pay standardization in Sweden probably helped destroy many smaller enterprises.

[45] Cameron, "Social Democracy, Corporatism, Labour Quiescence, and the Representation of Economic Interests in Advanced Capitalist Society," in John H. Goldthorpe, ed., *Order and Conflict in Contemporary Capitalism* (New York: Oxford University Press, 1984), 163–64, 177. Cameron does not explicitly cite Ingham there but relies on Ingham elsewhere to explain centralization and its consequences; see Cameron (fn. 41), 1256–57.

[46] Similar arguments are made by Nils Elvander, *Skandinavisk arbetarrörelse* (Stockholm: Liberförlag, 1980), 312, 320, 333; and Andrew Martin, "Wages, Profits, and Investment in

There is much intuitive appeal to this idea and possibly an element of truth. However, even strong corporatist unions have never been able to deliver solid or stable electoral and parliamentary majorities to left-wing parties and therefore to overcome what Adam Przeworski identifies as social democracy's dilemma in the struggle over votes—to go the route of radicalism with its attendant loss of moderate, often white-collar voters or to choose moderation and loss of radical, often blue-collar supporters.[47] Therefore, as Francis Castles demonstrates well, explanation of Social Democratic parliamentary domination must also consider the divisions and weaknesses of the parties to the right of Social Democrats.[48]

But even political division of the Right, characteristic of Scandinavia, probably does not adequately complete the explanation, because it fails to answer a nagging question posed by the concept of "relative power resources." If Swedish and Danish employers were as organizationally strong as depicted here, why were they unable to use that power to generate more frequent majority parliamentary coalitions against minority Social Democrats? Why with such resources did they fail to indoctrinate the electorate against Social Democratic ideas and to design strategies to unite conservative parties against the Left?

In part, they may simply have been unwilling to use their resources against the Left. The period of cross-class realignment in the 1930s, after which Social Democrats entered upon long periods of uninterrupted left-wing parliamentary control, was also a time when employers mostly withdrew from partisan politics and, with episodic exceptions, otherwise toned down their public propaganda against the Left. Instead of working primarily with and through rightist parties to win political influence, they bypassed the partisan route and dealt directly, confidently, and at times aggressively with Social Democratic governments.[49]

The concept of relative organizational resources in the hands of centralized labor confederations is therefore inadequate to the task of explaining the Left's political domination. Indeed, the historical record shows that it was employers' *strength*, not weakness, that helped bring about peak-level centralization at the same time the coalition was forged (at least in Sweden)—and then on distributional terms favorable to em-

Sweden," in Leon N. Lindberg and Charles S. Maier, eds., *The Politics of Inflation and Economic Stagnation* (Washington D.C.: Brookings Institution, 1985), 424–25.

[47] See especially Przeworski, "Social Democracy as a Historical Phenomenon," *New Left Review* 122 (July–August 1980), 27–58; and idem, "Party Strategy, Class Organization, and Individual Voting," in Przeworski (fn. 2), 99–132.

[48] Castles (fn. 4), 3–45.

[49] See, e.g., Sven-Anders Söderpalm, *Direktörsklubben* (The directors' club) (Lund: Rabén & Sjögren, 1976).

ployers. Employer satisfaction with unions' behavior in collective bargaining can also help account for high levels of union membership, which along with centralization is ironically one of Cameron's indicators of labor strength relative to capital.[50] One cannot account for the fact that Social Democrats held control so securely for so many years without recognizing the contentment of employers and their consequent disinterest in mobilizing voter and party sentiments against the Social Democrats. Had it been otherwise, they would have sought more energetically to merge liberals and conservatives and entice the Agrarian Party (later the Center Party) into alternative coalitions with them.[51] That employers were quite satisfied with Social Democratic governments can be seen in the apparent obsolescence of the lockout—as well as the strike—following the Social Democrats' rise to power in the 1930s.

Does Social Democratic Power Make the Strike Obsolete?

As many since Ingham have observed, the labor peace that ensued in Denmark and Sweden following the 1930s is remarkable, especially in light of the extraordinarily high levels of conflict earlier on in the Scandinavian countries. The balance of class power school of thought on the political economy of social democracy has a ready explanation, offered by Korpi and Shalev, Cameron, and others.[52] This argument suggests that both centralization and parliamentary control gave to the Left power that made the strike obsolete. Douglass Hibbs takes this argument the furthest, arguing that the 1930s brought "a massive shift of political power away from business interests and their middle-class allies to the 'organized working class.'" The benefits gained thereby in the form of "collective consumption and distribution" made the strike obsolete as an instrument for achieving distributional and political ends.[53]

The focus on employers as powerful institution builders and the view of Social Democratic rule as a cross-class alliance with capital in this study casts doubt on the importance of labor's political strength in the "withering away of the strike" in Scandinavia. James Fulcher is one of the few to recognize that in Sweden at least the strike had already begun

[50] On the importance of employer recognition for union membership levels, see Clegg (fn. 5).

[51] That they could have done so is suggested by Leif Lewin's intriguing analysis of the fragility of the initial red-green alliance and the potential ability of Agrarians to be bought off by the Right as well as the Left; see Lewin, *Ideologi och strategi* (Stockholm: Norstedt, 1984).

[52] Korpi and Shalev (fn. 1, 1979 and 1980); Cameron (fn. 45, 1984).

[53] Hibbs, "On the Political Economy of Long-Run Trends in Strike Activity," *British Journal of Political Science* 8 (April 1978), 153–75.

to decline in the 1920s, possibly due to the the new labor court.[54] Mo
puzzling in the usual discussion of the statistical findings is the consiste
failure to distinguish between the strike and the lockout. In their expla
nations of the decline of the strike in Scandinavia, all studies measur
things like "days lost in disputes," a large portion of which—in som
high-conflict years virtually all of them—were due to lockouts. In 192
for example, 98.7 percent(!) of Danish workers directly involved in dis
putes were locked out, and the proportion of days lost due to lockout
would have been similar. Because of the nature of officially reporte
statistics on work stoppages, Hibbs's composite measure of "strike vol
ume"—for Denmark at least—included man-days lost due to lockout
as well as strikes.[55]

Failure to distinguish between strikes and lockouts yields a distorte
picture of the decline of industrial conflict in Scandinavia, and the ques
tion of why the lockout vanished faster than the strike is left unad
dressed. From 1919 to 1932 in Denmark, on average 36 percent of work
ers directly involved in stoppages in any year were locked out rather tha
on strike. Between 1933 and 1955—after the Kanslergade Agreemen
and the introduction of compulsory arbitration by the Social Demo
crats—the figure plunges to about 5 percent. In Sweden, from 1919 to
1938 (Saltsjöbaden), on average each year 29 percent of workers in dis
putes were locked out or in "mixed disputes" involving both strikes and
lockouts. From 1939 to 1949 that figure slips to below 2 percent.[56]

Since the lockout is usually called in defense against strikes initiated
by unions, one might argue that its decline is simply a reflexive conse
quence of the decline of the strike. But this argument misses a key fact
about the lockout in Denmark and Sweden—that it was both defensive

[54] Fulcher (fn. 1). But the court was not established until 1928, after the decline set in. Also,
like Stephens and Luebbert, Fulcher errs in suggesting that the 1938 agreement followed a
shift in employer strategy in response to Social Democratic power (pp. 237–42).

[55] Figures from Det Statistiske Departement, *Statistisk Aarbog for Danmark*, various years.
Denmark's official statistics do not break down lost man-days as do those from Sweden.
Korpi and Shalev (fn. 1, 1980) do not discuss their sources, but the deficiencies of official data
make it unlikely that they excluded locked-out workers in their measure of "strike involve
ment." See also Ingham (fn. 41), 30. All seem to have taken their cue from Arthur M. Ross
and Paul T. Hartman, who deliberately combined strikes and lockouts even when they were
reported separately; see Ross and Hartman, *Changing Patterns of Industrial Conflict* (New
York: Wiley, 1960), 184. The same holds for the comparative analysis of strikes in Edward
Shorter and Charles Tilly, *Strikes in France 1830–1968* (London: Cambridge University Press,
1974), 306–34.

[56] After 1955 in Denmark and 1949 in Sweden, official statistics stop reporting lockouts.
With the virtual disappearance of lockouts from 1937 onward in Denmark (when Social
Democrats began using compulsory arbitration against the unions rather than against em
ployers) and from 1939 onward in Sweden, the respective statistical bureaus stopped report
ing them. Figures from Statistiska Centralbyrån, *Statistisk årsbok för Sverige*, various years;
and Det Statistiske Departement, *Statistisk Aarbog for Danmark*, various years.

nd offensive simultaneously. It was useful in the first instance to block ,orkers' immediate demands and whipsaw tactics by draining central :rike funds. Empty strike funds brought at least temporary respite in ɔllowing years. It was also used offensively in the 1920s to impose wage :ductions. Furthermore, it was instrumental in forcing upon unions, gainst the will of many among them, the kind of centralized authority eeded to control interindustry pay shifts. The September Agreement of 899 in Denmark was pushed on the unions with an encompassing lock-ut offensive. The 1938 Saltsjöbaden Agreement in Sweden was also ne-otiated from a position of strength: employers were genuinely surprised 1 1933 by the success of their lockout threat in spurring LO and the Social)emocrats into action against construction unions.

One could also argue that the rapid decline of the lockout was a sign f labor's new political power, which deterred employers from using the ɔckout. Jackson and Sisson argue plausibly however that the relative ,eakness of the Swedish Right early in the century, compared with else-,here in Europe, probably encouraged employers to rely on their own evices rather than on the state and therefore to become as organized nd militant as they did. Why then would the further decline of the .ight not make employers even more militant? In fact, employers in weden continued to threaten sweeping lockouts as late as the 1950s. 'hey did not have to use them, however, for the Social Democratic gov-rnment supported employers' objectives at the time, including even fur-ıer centralization of LO and the bargaining process. Swedish Social)emocrats never used their legislative power to challenge the legal right f employers to use the sympathy or any other kind of lockout, whereas ıbor in other countries fiercely opposes the right of employers to use the ɔckout. LO after all had promised to respect employers' right to the ,mpathetic lockout in the 1938 agreement.

With these considerations in mind, one is tempted to turn Korpi's and Iibbs's argument on its head and say that employers ceased using the ɔckout because, with the help of Social Democratic governments, they :hieved distributional objectives that they had previously tried (unsuc-:ssfully) to accomplish on their own. In other words, the political power ained by employer confederations in their cross-class coalitions with So-ial Democrats, unions, and farmers also caused them to lay down their ıost powerful weapon, one of the main purposes of which had been to alt the spread of wage increases and militancy from one sector to an-ther. By allying with Social Democrats, employers in traded-goods sec-ɔrs got the help they needed to stabilize the intersectoral pay structure, 'hich they had only controlled with repeated and costly shows of force.

The Danish Kanslergade Agreement of 1933, which introduced stat tory pay setting based on compulsory arbitration, was initially used forestall a lockout; soon afterward Social Democratic intervention w applied *instead* of lockouts.

In conclusion, the decline of industrial conflict in Scandinavia is r well explained by the idea of the growing power of the labor moveme at the expense of capital. Using the lockout, employers had partially su ceeded on their own in centralizing union control and reducing wo stoppages. Before the First World War and during the interwar yea employers in Denmark, unlike those in Sweden, asserted their centr ized power without help from Social Democrats, with the possible resu that conflict levels in Denmark were substantially lower.[57] Reduction conflict levels after the 1930s occurred in part because the lockout rath than the strike had essentially become outmoded as an alternative to p litical influence and in part because the strike had become more risky f the militant unions left out of the alliance. Social Democratic gover ments, through compulsory arbitration in Denmark and threats of t same in Sweden, helped create centralized discipline along lines desir by employers.

Reconceptualizing Labor's Power: Market Position and Class Organization

This critique has sought to show that labor's organizational structu and parliamentary strength are not sources of power against capital. B focusing on these things, the balance of class power literature has mi understood but mostly neglected the role of markets. The convention focus on late industrialization and industrial structure has misled man away from the possibility that Scandinavia's past exceptionalism in lab politics is better accounted for by relatively tight labor markets and im mobile capital vulnerable to international competition. Employers re acted to labor's associated advantages by building their extraordinar lockout arsenal. In contrast, early-twentieth-century waves of immigra tion favored American employers (to the consternation of Swedish em ployers) with large pools of surplus labor; later on American employe enjoyed greater geographic mobility within and outside their larger an more insular home markets. Early in the century U.S. employers rou tinely and effectively used strikebreakers; after the 1930s they often sim ply moved to escape labor's regionally limited legal and political protec

[57] Between 1903 and 1935 the loss of man-days due to disputes relative to the number o organized workers in Denmark was greater than in Sweden in only five different years; i Norway, in only four different years. See Galenson (fn. 9), 192.

ion. Thus, the multiemployer lockout as a weapon to control and shape abor played no significant role in American industrial relations.[58]

Hence, markets in Scandinavia gave workers and unions enormous verall advantages, including the power to constrain the range of strate-ic choices employers could make. They did not however give union eaders the instrumental power to dictate the shape of industrial organi-ations on terms they favored. (That ability seems to have been more haracteristic of organized employers.) It therefore does not follow from he notion that labor's power derives from markets as well as organiza-ion and politics that Danish and Swedish workers could simply parlay heir overall market advantages directly into organizational or political ower against capital. The reality is much more complicated: although ne labor group's distinct market position may have empowered it gainst capital, the same advantage may have activated other groups to orm effective organizational and political alliances *with* capital.

So the case seemed to be in Denmark and Sweden, where union roups in a strong market position suffered material, organizational, and olitical losses as a result of the cross-class alliances of the 1930s. Here strong market position" refers to the relative ability through indepen-ent action—strikes or price increases—to defend or raise one's share of actor income. The self-identified losers in the centralization of pay set-ng were workers in construction (especially unskilled laborers in Den-nark) and in printing industries whose insulation from international narket discipline placed them in a relatively advantageous market posi-on. Their strong position was ultimately neutralized by the power of entralized actors intent on restraining or reducing their relative wages vith organizational and political force. The self-identified and real win-ers were employers in a *weak* market position—manufacturers of raded goods vulnerable to permanent loss of market shares if they re-ponded with higher prices to increasing wage and other costs origi-ating from sectors like construction. They asserted control within their nulti-industry confederations at the expense of less militant and orga-ized employers in construction and then used their confederations to heck the spread of wage pressures and rank-and-file militancy across ectoral lines.

Metalworkers placed in a weak market position by harsh international ompetition gained organizational power within their confederations nd secured greater employment security if not higher wages from the hanges in pay distribution resulting from centralization. Their union eaders stood to gain from neutralizing the discontent among the rank

[58] Only recently have Americans witnessed such a lockout—in that most American of all dustries, baseball.

and file generated by disturbances in the intersectoral and interoccupational pay structure, that is, when construction wages in Sweden rose and metalworkers' relative wages (and due to high rents, purchasing power) fell or when separately organized unskilled workers in Danish construction helped their counterparts close skill differentials in manufacturing.

Even if they lost little in material terms, construction employers in a relatively strong market position lost organizational dominance, political influence, and even entrepreneurial autonomy in the new institutional and political order built upon centralization. Danish manufacturers rather than builders benefited materially from the Social Democrats' interventionistic exchange and trade policies; builders would perhaps have been better off with the 20 percent wage reduction instead. While aggressive export employers in Sweden frightened Social Democrats during the early postwar period from asserting political control over investment and production, they actually encouraged government regulation of housing construction (countercyclical and locational) to meet manufacturers' manpower needs. In any event builders' basic economic position was probably enhanced on the demand side by the extensive housing programs initiated by Social Democrats during the depression. Indeed, the extra employment opportunities created in both countries by the crisis programs probably also helped compensate construction workers for their organizational and distributional defeats.

In sum, those groups among both capital and labor that had enjoyed a strong market position were more directly disadvantaged by centralization and the wider sociopolitical alliances of the 1930s; by contrast, those with weaker market power put together the alliance and gained materially and/or organizationally. The winners within both capital and labor used their respective coercive power—the sympathy lockout and state intervention (or threats thereof)—to discipline the losers. The cross-class realignment of the 1930s thus equalized across sectors the power applied in regulation of the intersectoral income structure. In shaping the new order, the collaborative deployment of organizational and political resources across class lines by groups sharing similar market positions proved more important than were resources generated and applied according to the logic of class polarization.

Conclusion: Rethinking Social Democratic Alliances

The cross-class political realignment that brought Social Democrats in Denmark and Sweden to power and helped them stay there included

political alliances of union and employer groups behind centralized
trol of industrial relations and the intersectoral structure of pay.
thin both capital and labor, conflicts between traded-goods sectors
: engineering and sheltered home-market sectors like construction
tivated the formation of cross-class alliances for control of intraclass
npetitors and opponents. Employer lockouts and Social Democratic
rvention pushed in the same, not opposite, directions and for similar
poses. Social Democratic power substituted well for the lockout, and
refore the loss of working days due to industrial strife (including mas-
lockouts) subsided dramatically. The limited evidence available for
rway suggests the same factors help explain centralization, Social
nocratic rule, and industrial peace beginning in the 1930s.[59]

Centralization and Social Democratic governments promised to hold
an wages down and thereby limit rural wage costs and the loss of
or to industry. Thus the cross-class industrial alliance favoring cen-
ization in Denmark (primarily at the expense of unskilled laborers)
bably facilitated and stabilized the parliamentary coalition between
agrarians and Social Democrats. In a similar horse trade of agrarian
protection for urban employment, Swedish farmers made their support
for creation of new jobs in housing construction conditional on control
over militant construction workers and on limiting their high wages.
Further, though Swedish building workers lost in terms of bargaining
autonomy and relative wages, the depression-era construction program
promised more jobs. The parliamentary alliance therefore helped repair
divisions that had been created by the industrial-level alliance.

In conclusion, the cross-class alliance process that facilitated and sus-
tained Social Democratic rule in Denmark and Sweden was far more
complex, multilayered, and even time-consuming than has commonly
been depicted. Viewing centralization of industrial relations as a step in
the building of this alliance requires abandoning the common notion that
capital was weak or passive relative to labor (or "the state" for that mat-
ter), and hence not a "social foundation" of the 1930s alignment. Cen-

[59] The 1935 agreement between the labor and employers' confederations (similar to the 1899 Danish and 1938 Swedish agreements) inaugurated Norway's "solidaristic wage policy" of holding back pay in construction, in consonance with employers' lockouts and legal strategies; Walter Galenson, *Labor in Norway* (New York: Russel and Russel, 1970), 80–81, 175, 245. Social Democrats and the metalworkers' union joined employers in support of centralized control over militant building tradesmen (who bolted from the trade union confederation), especially high-pay masons in Oslo. See Svein Dahl, "Norsk Arbeidsgiverforening 1927/28: Talbaketog og revurdering" (The Norwegian employers confederation 1927/28: Retreat and reevaluation), *Historisk tidsskrift* 1 (1981), 1–25; and Erling Petersen, *Norsk arbeidsgiverforening, 1900–1950* (The Norwegian employers' confederation, 1900–1950) (Oslo: NAF, 1950), 532–48.

tralization did not, as is often supposed, give the Left added power to assert itself in parliamentary politics at the expense of capital. Nor did parliamentary success in alliance with farmers give labor the extra boost of power that it needed to centralize and therefore unify labor at the expense of employer power. Both of these ideas mistakenly attribute major institutional and political changes to a change in the balance of power between the undifferentiated and antagonistic blocs "capital" and "labor." By ignoring employer organization, strategy, and collective action *that led to centralization*, these notions distort our understanding of labor's limits—and opportunities—in capitalist democracies.[60]

[60] See also Swenson (fn. 3).

SPECIFIC FACTORS, CAPITAL MARKETS, PORTFOLIO DIVERSIFICATION, AND FREE TRADE

Domestic Determinants of the Repeal of the Corn Laws

By CHERYL SCHONHARDT-BAILEY*

TWO strands of thought relating to Britain's historic move to free trade in 1846 offer contradictory interpretations of the underlying lomestic economic and political forces at work. On the one hand, the ℓicardo-Viner specific factors model implies that owners of two facors—land and capital—stood diametrically opposed to one another on he issue of free trade. In the end, according to this view, capital gained he upper hand as seen in the 1846 repeal of the Corn Laws. On the other ıand, studies in the economic history literature posit that the economic nterests of these two groups of factor owners were not mutually excluive; rather, their interests overlapped as a result of rapid economic hanges in the 1830s that intensified landowner diversification into non-ǥricultural ventures. The implication of the latter approach is that landwners as a group came to be divided between undiversified landowners, 'hose economic interests remained tied to agriculture, and diversified ndowners, whose interests in agriculture had lessened while their in'rests in nonagricultural sectors had increased. Hence, the undiversified roup remained the only true "losers" from free trade in grain; members f the diversified group, by contrast, stood to gain or simply became ıdifferent to free trade.

* An earlier version of this paper was presented at the 1989 Joint Annual Convention of e British International Studies Association and the International Studies Association and the 1989 All-U.C. Group in Economic History Conference. Comments from John Conyare, Jeff Frieden, Barbara Geddes, Arye Hillman, David Lake, Peter Lindert, Tim Mccown, and Daniel Verdier were very helpful. My appreciation to Andrew Bailey for painsking editorial assistance, to Dan Steinberg for statistical advice, and to Craig Schonhardt r assistance with the tables. Support from the UCLA Center for International and Strategic ffairs and the UCLA Department of Political Science made possible the research for this per.

Of course, inherent to both arguments is the assumption that membe
of Parliament voted according to the economic interests of their consti
uents. Elsewhere, I have demonstrated the validity of this assumptio[1]
The scope and intent of this paper, then, is twofold. First, it discusses t
specific factors model and its inadequacies as applied to nineteent
century Britain, and it introduces two modifications that allow the mod
to account for the anomalies we observe in the British case. That is,
alter two key assumptions of the specific factors model by introduci
the concepts of portfolio diversification and investment capital flows in
the framework of factor specificity in the short run. The predicted poli
ical consequence of this modified version of the model is that diversi
cation lessened widespread protectionist sentiment among the lande
elite. In the context of a broader equilibrium model of trade policy li
eralization, it disrupted the initial protectionist equilibrium by decrea
ing the political benefits that accrued to M.P.'s from voting in favor
protection.[2] Portfolio diversification alone did not give rise to free tra
in Great Britain; but rather it was one of several factors that shifted th
political cost/benefit ratio of M.P.'s in favor of free trade.

Second, this paper extends the concept of portfolio diversification, a
presented in the economic history literature, to the realm of policy
making, using a statistical examination of the correlation between dimin
ishing protectionist sentiment—through diversification—and the fre
trade policy outcome.

The Specific Factors Model and the Long- and Short-Run Dimensions of Capital Formation

The Ricardo-Viner specific factors model in neoclassical trade theory defines some factor inputs as industry specific and others as intersectorally mobile. Industry-specific factors give rise to industry-specific interests that may either favor or oppose protection, depending upon the form of trade competition facing the industry. Owners of factors specific to the domestic import-competing industry gain from protection (via a relatively higher price obtained from the industry good), while owners of factors specific to the export sector lose (similarly, via the relatively lower price of the industry good). The preference of the mobile factor owner (for example, labor) is ambiguous; that is, although he can move into the

[1] Cheryl Schonhardt-Bailey, "A Model of Trade Policy Liberalization: Looking Inside the British 'Hegemon' of the Nineteenth Century" (Ph.D. diss., University of California, Los Angeles, 1991), chap. 3.

[2] Ibid., chap. 1.

rotected sector, his welfare remains contingent upon his unique consumption preferences. It follows that owners of factors specific to a paricular industry will tend to seek protection for their industry and oppose rotection for any other industry, whereas mobile factor owners will renain largely inactive.

Standard trade theory texts applying the specific factors model to Britin and the repeal of the Corn Laws posit (1) that landowners (that is, wners of factors specific to the import-competing industry) and capitalsts (that is, owners of factors specific to the export sector) deriving inomes from undiversified claims to factors opposed each other and (2) hat after 1832 industrialists had taken over Parliament, thereby engiieering the shift toward free trade.[3] The standard application of the speific factors model to the British case is correct in assuming that factors emained industry specific and that trade policy had the effect of redisributing income, thereby creating unambiguous winners and losers. owever, the two points posited are incorrect.

Taking the second point first, while the 1832 Reform had extended ffrage to many middle-class manufacturers,[4] it did not, by 1846, subantially alter the composition of Parliament. The Parliament of 1841– remained firmly under the control of landowners. By one estimate, out 80 percent of Parliament consisted of the landowning aristocracy d gentry.[5] Certainly if landowners as a group in Parliament wished to lt the repeal legislation, they had the votes to do so.

It is, however, the first problematic aspect to which this paper is dicted. The specific factors model implicitly assumes that (1) undiversid holdings of factors prevail (that is, each person owns only one, or imarily only one, factor of production), and (2) capital as a factor of oduction refers to the stock of fixed capital available for productive irposes. Extensions to neoclassical trade theory have at times relaxed e first assumption but have generally retained the second. Mayer, for stance, constructs an equilibrium tariff model that allows each person own more than one factor and that allows factor shares to vary among ople.[6] Each factor owner has an optimal tariff rate, the value of which determined by the individual's factor ownership. The equilibrium tar-

[3] This application appears in Richard E. Caves and Ronald W. Jones, *World Trade and yments* (Boston: Little, Brown, 1985).

[4] Before 1832 a significant number of manufacturers would have had sufficient property alifications to vote; they would have done so, however, in constituencies where they contuted a small percentage of the electorate.

[5] W. O. Aydelotte, "The Country Gentlemen and the Repeal of the Corn Laws," *English storical Review* 82, no. 322 (January 1967), 51.

[6] Wolfgang Mayer, "Endogenous Tariff Formation," *American Economic Review* 74, no. 4 84), 970–85.

iff thus hinges upon the underlying factor-ownership distribution, an
is the median factor owner's optimal tariff that determines the ac
tariff rate. Mayer applies the idea of diversified holdings both to the lo
run Heckscher-Ohlin model and its resulting Stolper-Samuelson Th
rem and to the short-run specific factors model. His intent is no
bridge the two models through the mechanism of portfolio diversif
tion; rather, it is to explore the implications of his modified assumpti
for the long-run adjustment of tariffs to changes in the distributior
factor ownership, voting costs, and voter eligibility rules and for
short-run attempts of small minorities of factor owners—under majo
voting—to obtain tariff protection for their industries.

In contrast to Mayer's application of diversified holdings, this pa
alters part of the assumption of undiversified holdings in the spec
factors model and considers capital in two forms—as stocks and flo
With regard to diversified holdings, I argue that income accruing fr
the ownership of a specific factor (for example, capital originating fr
the productive use of land) could be invested in other, more profita
sectors of the economy.[7] As a consequence, the allocation of capital flc
differed markedly from the allocation of capital stocks. Moreover, t
flow of investment capital from one sector to another was facilitated
the newly emerging market in long-term capital, as evidenced by
rapid growth in British stock market activity during the second quar
of the nineteenth century.

Theoretically, the beginnings of a capital market can be interpreted
the start of a transition from a state in which capital is fixed to one
which it becomes more mobile between sectors. In Britain capital flo
responding to unequalized returns, shifted into higher yield industr
sectors of the economy, thereby altering the ownership distribution
capital stock. (Hence, a corollary argument may be given for capital m
ket creation; that is, the more the allocation of flows diverges from
allocation of stocks, the greater the need for a capital market.) It shou
be acknowledged, however, that other forces may have come into pl
to delay the emergence of a long-run equilibrium for returns to inve
ment across the different sectors. In the short run the new free-tra
policy may have added to the existing inequality of returns by rewardi
the owners of industrial capital (unless the capital market had previou
fully taken into account the returns from a policy change—an unlik

[7] For an excellent discussion of nineteenth-century British investors' responsiveness to fl
tuations in investment opportunities, see R. C. Michie, *Money, Mania and Markets: Investm
Company Formation and the Stock Exchange in Nineteenth-Century Scotland* (Edinburgh: Jo
Donald, 1981).

:nt in such a nascent market). Thus, in the short run capital flows ıld, if sufficient, have generated a trade policy shift that created more :quality of returns, not equalization, even though in the very long run :urns should have equalized. Nonetheless, the essential point is that : new ownership distribution created economic and political incentives · a policy change that favored owners of industrial stock (for example, nove to remove protection for agriculture). This said, the contribution this paper is quite modest: it introduces portfolio diversification in ıeteenth-century Britain into a medium-run model with a "dynamic" mponent, that is, capital flows that reflect divergent returns in industry d agriculture.

More to the point, however, portfolio diversification significantly al-'s the political interpretation originally posited by the specific factors ›del. Protective tariffs continued to be ardently favored or opposed by : owners of factors tied to specific industries.[8] However, the policy nce of factor owners whose returns could be reallocated elsewhere :ame ambiguous. This was so especially where these returns (in the ·m of investment capital) found their most profitable outlet in indus-cs specific to other factor inputs. The interests of these individuals ›uld have derived from each person's unique portfolio—that is, from : extent of diversified holdings or, interpreted strictly, from the dis-ınted expected future income stream from holdings under each policy ernative (the latter clearly more difficult to gauge than the former).[9] nply put, as landholders diversified more extensively, their interests her began to resemble more closely those of industrialists favoring free de or became less sharply defined, thus bordering on indifference. ıly those individuals who could not diversify out of agriculture (for ımple, tenant farmers and landowners on the verge of insolvency) uld remain staunch advocates of protectionism.[10] Since the specific tors model misses the importance of a growing British capital market, fails to recognize that the support for protectionism was diffusing

These were individuals who were under some compulsion—perhaps impending insol-cy—to reinvest flows in a factor that yielded less than (an) other factor(s).

In an ideal world one would not engage a static model but rather would look to a dy-ıic one that incorporated future expected returns.

'It should be mentioned that undiversified farmers were enticed by the government to uiesce to repeal, as evidenced most clearly in the "drainage loan" portion of the repeal slation; see D. C. Moore, "The Corn Laws and High Farming," *Economic History Review*, ser., 18, no. 3 (1965), 544–61. Nevertheless, many "marginal" farmers (i.e., grain farmers ble to increase production through advanced farming methods or unable to switch to :r forms of farming) left Britain in the wake of repeal hoping to gain better returns for r capital investments in the U.S.; see William E. Van Vugt, "Running from Ruin?: The igration of British Farmers to the U.S.A. in the Wake of the Repeal of the Corn Laws," *nomic History Review*, 2d ser., 41, no. 3 (1988), 411–28.

within the ranks of those who supposedly were its leading advocates. The division between owners of factors and their supposed conflict of interests, then, is much fuzzier than the model implies.

The Evidence

Evidence for the portfolio diversification hypothesis is presented below in two forms—empirical and statistical. The former draws from the economic history literature and offers a compelling account of the progression of diversification, especially its escalation from the late 1820s to the 1840s, a period of rapid growth in the newly emerging stocks and shares market. The latter attempts to establish a link between growing diversification and M.P. voting behavior on the question of free trade in grain.

Diversification in Economic History

Diversification of portfolios by landowners was by no means a new development in the nineteenth century. Beginning in the late sixteenth and early seventeenth centuries, both the older aristocracy and the rising gentry invested extensively in nonagricultural activities, including minerals and mining (coal, lead, steel, salt, alum, and so forth), urban development, shipping, and joint-stock companies. Lawrence Stone explained that even though "industry was not the road to great riches," it provided to both peers and gentry an interesting diversion, and an extension of the taste for gambling. In his analysis of 158 peerage families during the period 1560–1639, Stone found that 25 percent of these profited by mining activities on their estates; 9 percent invested in fen drainage; 15 percent in developing London; 14 percent in shipping; and 63 percent in trading, colonial, and industrial concerns. Moreover, he found the older nobility no less willing to invest in nonagricultural activities than was the newer nobility.[11]

J. H. Plumb extended the notion of diversification to the eighteenth century. According to Plumb, cheap water transport (provided through canal ventures) allowed greater diversification of portfolios among the gentry.[12] The opening of rivers and coastal traffic both enhanced agricultural production by creating new metropolitan markets for agricultural goods and created new outlets for enterprises in timber sales, gravel, and minerals.

[11] Stone, *The Crisis of the Aristocracy 1556–1641* (Oxford: Oxford University Press, 1965), 377–83.

[12] Plumb, *The Growth of Political Stability in England, 1675–1725* (London: Macmillan, 1967), 5.

Consequently, by the early nineteenth century some landowners were t least marginally involved in industrial activities. The obvious question s then, why, given some degree of diversification, did landowners sup- ort agricultural protection as long as they did? It may be that (1) land- wners diversified but the returns on their nonagricultural income were nconsequential relative to their agricultural income, (2) landowners' in- estment in nonagricultural sectors was small relative to their agricul- ural holdings, or (3) numbers of landowners actually diversifying re- nained few relative to those not diversifying. A fourth and quite lifferent explanation is that the type of industrialization in the eigh- eenth century differed from that in the nineteenth; whereas the former vas led by domestic demand, the latter was led by exports. Thus, land- wners diversifying into industry in the eighteenth century would likely ave retained their preference for agricultural protection, whereas land- wners similarly diversifying in the nineteenth century would not. In ny case, two critical economic changes in the second quarter of the nine- eenth century—the expansion of exports in mining and heavy industry nd the development of the capital market—marked that period as a urning point. There was an increase in (1) returns to diversified hold- ngs, (2) the size of nonagricultural holdings, and (3) the numbers of andowners diversifying. These changes, moreover, transformed the pro- ess of investing outside of agriculture—all of which dampened land- wner support for agricultural protection. Interestingly, both changes eem to have been triggered by (or to have coincided with) the same vent—the beginning of the railway boom in the mid-1830s (peaking in 836–37 and again in 1844–46).

The first such sudden economic change was the expansion of exports n mining and heavy industry. While export growth and the railway ooms should actually be considered a two-way causal relationship,[13] this

[13] The literature explaining the various linkages between railways and the development of eavy industry in Britain is considerable. For instance, Llewellyn Woodward wrote that "the ailways were one of the results of progress in the iron industry and of the increased con- umption of coal brought about by the use of steam power. They were in turn the cause of a ast expansion in the metal trades and of a much greater demand for coal. Directly or indi- ectly they influenced the development of most industries in the country. [In addition] the ailway itself was an article of export; British contractors built lines in every continent and rganized companies to buy them"; see Woodward, *The Age of Reform, 1815–1870* (Oxford: xford University Press, 1962), 41. E. J. Hobsbawm found that the "immense [coal] industry, hough probably not expanding fast enough for really massive industrialization on the mod- rn scale, was sufficiently large to stimulate the basic invention which was to transform the apital-goods industries: the railway. For the mines not only required steam-engines in large uantities and of great power, but also required efficient means of transporting the great uantities of coal from coal-face to shaft and especially from pit-head to the point of ship- ent. The 'tramway' or 'railway' along which trucks ran was an obvious answer. . . . Tech- ologically the railway is the child of the mine, and especially the northern English coal

sudden export growth nevertheless was created in the very industries (coal, iron, steel) in which landowners had been longtime investors.[14] Although their portfolios already included such investments, the intensity and economic importance of these investments had increased with the growth of these industries.[15]

The second change was the development of the capital market. Although the beginnings of a formal market in capital may be traced to long-term borrowing by the state in the late seventeenth century, it was not until the early 1800s that the London Stock Exchange drafted a constitution and obtained its own building. The market of the eighteenth century was not a national market; rather, it was concentrated in London and largely confined to the trade of government securities.[16] From the beginning of the nineteenth century to the mid-1820s, very little share trading existed outside London. However, with the repeal in 1825 of the Bubble Act (which had required all companies to obtain a royal charter or an Act of Parliament and made it illegal for any broker to buy or sell

---

mine"; see Hobsbawm, *The Age of Revolution, 1789–1848* (New York: New American Library, 1962), 63. He added that "in the first two decades of the railways (1830-1850) the output of iron in Britain rose from 680,000 to 2,250,000 [tons]. . . . The output of coal between 1830 and 1850 also trebled from 15 million tons to 49 million tons. That dramatic rise was due primarily to the railway, for on average each mile of the line required 300 tons of iron merely for track" (p. 64).

B. R. Mitchell clearly sketches the correlation between rising demand in the iron and coal industries and the railway booms and, moreover, the subsequent royalties accruing to land owners. For example, Mitchell explains that royalties usually consisted of a fixed rent plus a royalty paid according to the amount of coal extracted. From the early nineteenth century to the 1870s, royalties averaged between 8% and 11% of coal sales, or 6d to 9d per ton, or £50 to £180 per acre; see Mitchell, *Economic Development of the British Coal Industry, 1800–1914* (Cambridge: Cambridge University Press, 1984), 251–58.

It should be noted that while recent British historiography has discounted the Rostowian notion of a sudden industrial "takeoff" period, such revisions have applied not to the heavy industry stage of industrialization beginning in the 1830s but rather to the period from roughly 1780 to 1830; see N. F. R. Crafts, *British Economic Growth during the Industrial Revolution* (Oxford: Oxford University Press, 1985). Other useful sources that address the linkages between railways and the development of heavy industry (not to mention the large literature simply on railway development) include P. L. Cottrell, *Industrial Finance, 1830–1914* (London: Methuen, 1980); F. M. L. Thompson, *English Landed Society in the Nineteenth Century* (London: Routledge and Kegan Paul, 1963); and G. N. von Tunzelmann, *Steam Power and British Industrialization to 1860* (Oxford: Oxford University Press, 1978).

[14] See Thompson (fn. 13) for a discussion of the mineral incomes of several great landowners (e.g., Dukes of Northumberland and Portland, the Earl of Carlisle, Lords Hastings and Rokeby, Sir Matthew White Ridley, and the Earl of Durham) and gentry (e.g., the Claytons, Crofts, Bates, Edens, Riddells, Wrightsons). He noted that "landowners certainly drew large and increasing incomes from coal, but these were predominantly and increasingly in the shape of royalty and wayleave rents" (p. 264). Some gentry received mineral income equal to half their total income.

[15] One might add that landowners would be far more likely to invest in industries (mostly heavy) that had land occupying a larger share of assets—i.e., coal, iron, railways—than such light industries as cotton textiles.

[16] See, e.g., Cottrell (fn. 13); and E. Victor Morgan and W. A. Thomas, *The Stock Exchange: Its History and Functions* (London: Elek Books, 1962).

shares in unchartered companies), share trading boomed as joint-stock companies—most notably insurance and banking companies—could now sell their shares on the stock exchange.[17] Further legislative restrictions on the privileges of incorporation were lifted in 1834, 1837, and 1844, thus contributing to the growth of trading in joint-stock companies.[18]

More importantly, after the successful opening of the Liverpool and Manchester line in 1830, confidence in railway stock surged, as evidenced by the "railway manias" of the mid-1830s and mid-1840s. After 1830 a company needed only to advertise in a railway journal that stocks were available to be flooded with applications. For example, in 1836 the New Gravesend Railway received 80,000 applications for the 30,000 available shares; the Great North of Scotland Railway, more than three times the number of its shares in 1845; the Direct Western Railway, 1,400,000 applications for its 120,000 shares; and the Direct London-Exeter, 400,000 for its 120,000 shares.[19]

Aside from later technological changes (telegraph, telephone), railway shares offered the single most important factor in the integration of Britain's emerging capital market—a common security that could be traded actively on more than one market. Without common securities, "localized imbalances in the supply of, or demand for, stocks and shares would continue to result in dramatic price fluctuations and an inability to meet requirements. With common securities, local price changes would generate an immediate flow from, or to, that centre so that all markets would rise and fall in line."[20]

Such common securities contributed especially to diminishing the previously localized nature of the provincial markets, as active trading between markets outside of London expanded throughout the 1830s and 1840s. Elsewhere I illustrate the rapid growth in provincial stockbroking and find that railway stocks propelled this growth.[21] Indeed, according

[17] Manufacturing and industrial enterprises were not as apt to benefit from the repeal of the Bubble Act since, unlike insurance companies whose ownership was corporate, these firms largely consisted of partnerships; see Michie (fn. 7), 62.

[18] See Morgan and Thomas (fn. 16), 125–31.

[19] Harold Pollins, "The Marketing of Railway Shares in the First Half of the Nineteenth Century," *Economic History Review*, 2d ser., 7, no. 2 (1954), 233.

[20] R. C. Michie, "The London Stock Exchange and the British Securities Market, 1850–1914," *Economic History Review* 2d ser., 38, no. 1 (1985), 68.

[21] Schonhardt-Bailey (fn. 1), chap. 4. Moreover, J. R. Killick and W. A. Thomas provide a listing of companies quoted on the Leeds, Liverpool, Manchester, and Newcastle exchanges that clearly illustrates the growing dominance of railway shares from 1837 to 1847; see Killick and Thomas, "The Provincial Stock Exchanges, 1830–1870, *Economic History Review*, 2d ser., 23, no. 1 (1970), 96–111.

to one estimate, railway construction accounted "almost exclusively" for the establishment of the provincial long-term capital market.[22]

Nor was the emerging capital market limited to the English provinces; as the onslaught of railway and joint-stock promotion swept across the country, it sparked similar share trading in Scotland. During the period 1830–40 the massive expansion in the Scottish coal and iron industry led to an expansion of mining and heavy industry, which in turn required substantial investment capital.[23] Concurrently, the number of investors benefiting directly from the output of the enterprise (a feature of the earlier localized nature of investment) decreased, while the number of investors seeking primarily to maximize the return on tradable investment rose. By the end of the 1830s a national share market had developed in Scotland, with insurance companies and railway shares experiencing the most rapid growth. The early 1840s brought economic prosperity and with it surplus funds seeking investment opportunities. Given the excess funds in circulation, banks lowered their interest rates,[24] and consequently, stocks became "the most eligible and favourite" form of investment. Moreover, during the early 1840s higher rates of return from railway shares markedly altered the structure of the Scottish investment economy by elevating the importance of railway capital at the expense of all other investment sectors.[25]

Just as railway shares provided a common security for the provincial stock exchanges, so too did they link the emerging Scottish exchanges with their English counterparts. In 1846, for instance, English shareholders were the source of 38 percent of the funds provided for Scottish railways by large investors.[26] This cross-investment was not one-sided, however, as Scots similarly sought stock in English railways.[27] Finally, in addition to the provincial and Scottish stock exchanges, the London Stock Exchange itself exhibited the speculative mania associated with the railways. From 1844 to 1846 the annual number of Railway Acts passed

[22] Killick and Thomas (fn. 21), 110. See also W. A. Thomas, *The Provincial Stock Exchanges* (London: Frank Cass, 1973).

[23] Michie (fn. 7) estimates the gross capital formation in Scottish iron manufacture to have increased from 1.8 million *pounds* in 1828 to 6.6 million *pounds* by 1840 (p. 51).

[24] The Bank of England lowered its rate from 6% in late 1839 to 3% in 1846, while in Scotland, the rate fell from 3.5% in 1842 to 2% in 1843; see John Clapham, *The Bank of England*, vol. 2 (Cambridge: Cambridge University Press, 1944); and Michie (fn. 7).

[25] Michie (fn. 7), 92, 117.

[26] Ibid., 117.

[27] For example, "the Liverpool and Manchester Railway had only 10 Scottish shareholders in 1838, but 29 by 1845, while the Grand Junction Railway had 4 in 1835 and 122 in 1845. Scots had invested only £2000 in the Newcastle and Carlisle Railway in 1838, but £9,300 in 1844, while they held £94,000 of the stock of the Great North of England Railway in 1845 compared to a mere £4,700 in 1838"; see Michie (fn. 7), 117–18.

increased from 48 to 270, and capital authorized increased from £17.8 million to £136 million; total railway capital quoted on the London Stock Exchange thereby increased from £72.3 million to £126.1 million.[28]

Stock market activity, however impressive its growth, must necessarily be joined to the interests of landowners to warrant the claims that (1) landowners were indeed diversifying into nonagricultural ventures such as railway shares and that (2) this diversification had potential political ramifications. In a presentation of the occupational breakdown of shareholders in various railway and utility companies, I demonstrate that esquires and gentlemen held a strong interest in such investments.[29] One historian has conservatively estimated the average percentage of gentlemen/landowners among subscribers for all railway companies floated in 1844–45 to be 18 pecent,[30] while others have estimated the figure for the period 1820–44 to be 28 percent.[31] In fact, many railway companies deliberately reserved a certain percentage of shares (say, 20 percent) for landowners. Landowners were considered valuable shareholders not only for their resources, but also because a company with a considerable number of landowners as shareholders could use this as clout to prevent present or future opposition to the railway line from other landowners.[32]

Landowner interest in railways stemmed not only from the ownership of shares but also from the high prices paid to them by the railway companies for their lands. Rather than face the possibility of legal battles over the value of land—costing four times the price of those purchases unopposed—railway promoters "bought off" landowners by paying them, on average, twice the current market value for their land. In fact, in 1844 the Board of Trade estimated that the excess land costs alone made Britain's railways £1,000 more expensive per mile than railways on the Continent.[33]

Evidence from the economic history literature thus clearly traces increasing diversification of the landed elite into mining and heavy industry, as well as railway development. This is not to suggest, however, that such diversification directly translated into their conversion to a free-

[28] Morgan and Thomas (fn. 16), 106.

[29] Schonhardt-Bailey (fn. 1), chap. 4, presents selected data from Killick and Thomas (fn. 21).

[30] S. A. Broadbridge, *Studies in Railway Expansion and the Capital Market in England, 1825–1873* (Guildford and London: Frank Cass, 1970), 144.

[31] G. R. Hawke and J. P. P. Higgins, "Transport and Social Overhead Capital," in Roderick Floud and Donald McCloskey, eds., *The Economic History of Britain since 1700* (Cambridge: Cambridge University Press, 1981).

[32] Pollins (fn. 19), 238.

[33] R. J. Irving, "The Capitalisation of Britain's Railways, 1830–1914," *Journal of Transport History*, 3d ser., 5, no. 1 (1984), 14–15.

trade policy stance. Coal, iron, and steel exports certainly became increasingly more important in the 1830s and 1840s, but so, too, did domestic consumption of these goods (especially for domestic railway construction). Additionally, even landowner interests in railway promotion could in some cases be linked to the transportation of agricultural goods to domestic markets rather than to the development of British export industries. Nevertheless, diversification into potentially (and in some regions of the country, predominantly) nonagricultural ventures allowed landowners to spread their investment risks among various sectors of the economy not directly benefiting from the expansion (or maintenance) of British agricultural production. As stated earlier, the actual interests of these individuals must be derived from each landowner's unique portfolio, that is, the diversification of holdings (as a proxy for the discounted expected future income stream from his holdings). Either, in the extreme, landowner interests began to resemble those of industrialists favoring free trade, or more moderately, their interests simply became less sharply defined, perhaps bordering on indifference.

Diversification and M.P. Voting Patterns on Free Trade

Previous studies that addressed the issue of diversification in the repeal of the Corn Laws generally focused on the individual interests of members of Parliament. A primary source of motivation for voting, according to this view, stemmed from the M.P.'s personal pecuniary interests. (Indeed, one historian commented that the reason attempts by the 1841-47 government to regulate railways were doomed was that M.P.'s—the majority of whom were shareholders in railways and one-seventh of whom were railways directors—vigorously defeated any "dangerously restrictive" proposals.)[34] The argument continues that M.P.'s were beginning to diversify their personal portfolios in the years prior to 1846: as they increasingly invested in the business and industrial sectors of the economy,[35] they personally had less to lose and more to gain from eliminating agrarian protectionism.

My test of this hypothesis revealed that a transformation in the incomes of members of Parliament was indeed occurring, though not at a

[34] Asa Briggs, *The Age of Improvement, 1783–1867* (London: Longman 1959), 340.

[35] For general references to this argument, see Barrington Moore, *Social Origins of Dictatorship and Democracy* (Boston: Beacon Press, 1966); and E. J. Hobsbawm, *Industry and Empire* (London: Weidenfeld and Nicolson, 1968). For the specific application of this argument to the M.P.'s of 1841–47, see W. O. Aydelotte, "The Business Interests of the Gentry in Parliament of 1841–1847," in G. Kitson Clark, *The Making of Victorian England* (London: Methvent, 1962), 290–307; and idem (fn. 5), 47–60. See also J. A. Thomas, "The House of Commons, 1832–1867: A Functional Analysis," *Economica*, no. 13 (1925); and idem, "The Repeal of the Corn Laws, 1846," *Economica*, no. 25 (1929).

rate or level sufficient to confirm the hypothesis.[36] Timothy McKeown tested a similar hypothesis. Although he found a systematic relationship between voting behavior and "personal pecuniary interests" of M.P.'s, he did not cite this as the critical factor in the abolition of the Corn Laws.[37]

One step beyond gauging the economic effect of portfolio diversification on members of Parliament entails examining its political effects, as channeled through diversification by their constituents. It is the argument of this paper that as individuals (and constituencies in the aggregate) invested returns from land into nonagricultural ventures such as industry and railways, political support for protectionism waned, and M.P.'s accordingly shifted their votes toward free trade.

Two approaches, differentiated by data type (individual versus aggregate), are adopted below for testing the relationship between diversification and M.P. voting behavior. Neither approach is ideal, but taken together they provide a reasonable test of the hypothesis. A better test of diversification would rely exclusively on individual-level data in order to capture the "capital as flows" element of the argument. That is, whereas aggregate-level data may capture shifts in the overall distribution of land and capital ownership (and consequently diversification at the aggregate level), the underlying assumption is that all individuals have undertaken exactly the average diversification of their personal portfolios by investing returns from one factor into another factor. Landowners may instead have moved wholly into the industrialists category; alternatively, the number of industrialists may have grown more quickly relative to landowners. As such, individual and aggregate-level data will provide different pictures of diversification. Individual-level diversification suggests that capital operates as flows and that industry-specific interests become less sharply defined in the political arena. Aggregate-level diversification may indeed reflect individual-level diversification, in which case the same argument holds; that is, capital "flows" and interests merge. Alternatively, diversification may occur only at the aggregate level, whereupon interests remain industry specific: landowners remain tied to agri-

[36] In an earlier phase of this project, I tested this hypothesis. In sum, I found that while unmistakable change in the economic interest composition of Parliament was occurring from 1832 onward, the change was quite gradual. The landowning interest was steadily declining while that of business was steadily rising, overtaking the former by the 1870s. I concluded that it seemed likely that it was not the M.P.'s own economic interest that motivated him, but that of his constituents. Transformations in the economic interests of constituents thus took several decades to be reflected in the composition of Parliament.

[37] Timothy J. McKeown, "The Politics of Corn Law Repeal Reconsidered" (Paper presented at the 1987 annual meeting of the American Political Science Association, revised December 1987).

cultural protection, and industrialists, tied to free trade, although the balance of interests becomes more evenly distributed.

There are drawbacks to relying solely on microlevel data: such data are rare for early-nineteenth-century Britain, and the available data are incomplete and inconsistent. Consequently, I test the diversification thesis using both individual and aggregate data. Positive and consistent findings at both levels of analysis would support the hypothesis that diversification was indeed occurring at the individual level and that the more aggregate-level statistics were an accurate reflection of this diversification.

Testing the Hypothesis

Two sources of data were used to test the hypothesis that M.P.'s who represented constituencies with greater diversification were more likely to vote in favor of free trade than were those representing less diversified agricultural constituencies. The first source—death duty registers[38]—provides data at the individual level. The years 1830 and 1850 were chosen to construct a random sample of 1 percent for each year (162 and 184 individuals, respectively).[39] (Elsewhere I describe the sampling technique in full.)[40]

The components derived from the registers include (1) stocks (that is, stocks or shares in railways, utility companies, canals, joint-stock companies, insurance companies; investments in a business or trade; and mining interests—only when listed separately from real estate), (2) gov-

[38] Barbara English, "Probate Valuations and the Death Duty Registers," *Bulletin of the Institute of Historical Research* 58, no. 135 (May 1984), argues that death duty registers offer more complete records of business holdings than do probates. For example, the registers list stocks and shares, industrial machinery, business interests, and residual estate. They offer more details as to the form in which wealth was held at death. Only since 1982 have the registers been opened to study by researchers for the period from 1796 to 1903. I did find, however, that for later years many registers remain closed, some until the year 2007.

Jane Cox describes and explains the contents, format, notations, and abbreviations relevant both to probates and to death duty registers; see Cox, *The Records of the Prerogative Court of Canterbury and the Death Duty Registers: A Provisional Guide* (Canterbury: Public Record Office, Prerogative Court of Canterbury, 1980).

[39] Since 1825 and 1846 constitute roughly the critical years "before" and "after" the emergence of capital markets and diversification, and since we may assume that an individual would most likely be at his financial peak not at his death but rather a few years prior, 1830 and 1850 appeared to be adequate choices for each time period sample.

[40] Schonhardt-Bailey (fn. 1), chap. 4. In brief, however, a mapping of both the 1830 and 1850 samples illustrates a fairly even geographic distribution, with of course the expected concentration around London. A comparison of the regional distribution with actual population figures for England in 1831 and 1851 suggests that the sample distributions appear to match the population figures relatively well. It should be noted, however, that in 1830 the Southeast (including London) and to a lesser extent the Midlands appear overrepresented, while the northern regions are underrepresented. For 1851 the regions fairly accurately correspond to the population figures.

ernment securities (consols, reduced annuities, etc.), (3) charges on real estate (including both absolute and annuity bequests charged on real estate holdings, and real estate sold whose value was then taxed), (4) cash bequests (from bank stock and monies from sources unspecified), and (5) residue (which in many cases included the value of real estate property when instructions were given for its sale—the proceeds of which were then subject to taxation). Unfortunately, real estate was not subject to death duty and subsequently was not valued until the early 1850s. Variable (3) combined with variable (5) provides the best proxy for estimates of real estate, given the clear overlap between the two.

An examination of stockholdings by region and county illustrates the geographic spread of capital market activity: whereas in 1830 stockholders appeared in only four counties (all concentrated in London and the South), in 1850 stockholders appeared in thirteen counties (spreading regionally from the Southeast and Southwest to include East Anglia, the Midlands, the Northeast and Northwest, and Wales). Moreover, the percentage of individuals owning stock increased from 3.7 percent in 1830 to 12.5 percent in 1850, also supporting the notion of an expanding capital market discussed earlier.

The second data source—income tax returns—also provides data on the extent of diversification, but at the constituency level (that is, borough and county) and for the years 1814 and 1856. Although 1814 and 1856 were not "ideal" years (ideal would have been 1825 and 1845), the lack of any income tax structure between 1815 and 1842 meant that 1814 was the latest possible year of the first period (excluding the unrepresentative final wartime tax year, 1815).[41] Further, 1856 was chosen since it was the first year after reinstitution of the income tax that the Parliamentary Papers offer a complete county and borough breakdown for all the tax schedules. The two years are similar in that both were war years and both operated under a virtually identical tax structure.[42]

Income tax returns consisted of four schedules: Schedule A (profits from the ownership of lands, farm buildings, houses, tenements); Schedule B (profits from the occupation of lands, houses, tenements—in short, profits from farming); Schedule D (profits from trade or manufacturing business, profession, employment, or vocation, and miscellaneous items such as foreign securities and possessions); and Schedule E (annuities,

[41] As it was the final year of the wartime income tax, returns in 1815 dropped sharply due to public resistance to the tax and less stringent collection efforts by the Tax Office; see Arthur Hope-Jones, *Income Tax in the Napoleonic Wars* (Cambridge: Cambridge University Press, 1939), 77, 109. Hence, 1815 was not considered an appropriate year to sample.

[42] Ibid., 121.

pensions, and stipends paid to holders of public office). An adjustment of the data[43] enabled me to divide Schedule A into two parts, A1 and A2, with the former comprising income from rural landownership and the latter comprising income from urban landownership.

To determine whether or not M.P.'s from constituencies with greater diversification were more likely to vote in favor of free trade, I constructed indexes of diversification for each sample. Individual scores were averaged to obtain one score for each county in the death duty register sample, whereas separate scores could be obtained for boroughs and counties in the income tax returns sample.[44] Indexes used for both samples are given in Table 1. The first term in the death duty register diversification index (stocks/sum of all Xs) denotes the relative diversifi-

TABLE 1
INDEXES OF DIVERSIFICATION

Death Duty Registers

$$\text{Diversification Index} = \frac{X1}{\text{Sum } x\text{'s}} - \frac{(X3 + X4)}{\text{Sum } x\text{'s}}$$

where
- X1 = stocks and shares
- X2 = government securities
- X3 = charges on real estate
- X4 = residue
- X5 = cash, bank stock, money from unspecified sources

Income Tax Returns

$$\text{Diversification Index (for counties)} = \frac{D}{\text{Sum }(A1,A2,B,D,E)} - \frac{A1}{\text{Sum }(A1,A2,B,D,E)}$$

$$\text{Diversification Index (for boroughs)} = \frac{(D + A2)}{\text{Sum }(A1,A2,B,D,E)} - \frac{A1}{\text{Sum }(A1,A2,B,D,E)}$$

where
- A1 = income from rural landownership
- A2 = income from urban landownership
- B = income from farming
- D = income from trade, manufacture, etc.
- E = income from public office

[43] Schonhardt-Bailey (fn. 1), Appendix 2.

[44] Further on I separate the votes of M.P.'s into those representing boroughs and those representing counties. For the death duty register sample, this means that, statistically speaking, both sets of M.P.'s are responding to the same county-average diversification score.

nto stocks and shares, whereas the second term (charges on real - residue/sum of all Xs)[45] is the proxy for agricultural holdings.[46] ome tax returns diversification index follows the same approach; 'oadens the scope of the first term to include all profits from trade nufacture, and for boroughs it also includes profits from urban ıership. The second term isolates rural landowners as much as . The diversification indexes range from −1 to +1; the higher e, the greater the diversification into stocks and shares (in the the death duty registers), or the greater the diversification into ıder array of nonagricultural ventures (in the case of the income rns).

ıal clarification regarding the diversification indexes should be ıed. Since the possible scores range from −1 (complete invest- . land) to +1 (complete investment in trade/industry/etc.), one ırmise that the "perfect," or best, score would be zero. This con- might be warranted if one's argument were founded upon a non- nal diversification index, that is, one in which industrialists in- in land created the same incentive for M.P.'s to vote for free trade ındowners investing in industry. Such an index would presum- ıuire one to obtain the absolute values of each score and hypoth- ıegative correlation between diversification and free-trade votes free trade = 1 and protectionism = 0). In contrast, the argument hat landowners were seeking higher returns by investing in in- –necessitates a directional index and hypothesizes a positive cor- between diversification and votes for free trade. The higher the ' a particular district, the greater the probability of its represen- [.P. voting for free trade.[47]

'eal estate estimate for this sample suffers from various coding difficulties. Real ıally listed as "charges thereupon" rather than as actual values, creates distortion to inflated values in some cases and to unreported values in others. It is possible two distortions may have nullified each other, but without better microlevel data of the distortions is unknown. Moreover, real estate included not only agricultural ›ut also commercial property, and sometimes mining interests as well. Finally, clude "residue" with real estate, not all the residue value could indeed be verified ate. The effect of these cross-cutting distortions in the estimate of real estate is, :nown.

'ing both terms of the equation was considered but dismissed, since the result was ed sensitivity to changes in income levels. In other words, given equal ratios of real ıon–real estate holdings, squaring the terms multiplied the effect of one's income ı wealthier individual would be predicted to have an exponentially stronger interest ree trade or protection for any given ratio of real estate to non–real estate holdings. ıared version, being less responsive to changes in income levels, has the advantage g to a greater degree the effects of diversification.

lly speaking, diversification need not demonstrate direction. Returns from factor :d not be channeled from land to capital; the flow could be reversed. The point is

Diversification indexes for both samples are divided into "level" variables and "difference" variables. The former are simply the diversification scores for each sample year (1830 and 1850 for the death duty registers, and 1814 and 1856 for the income tax returns), while the latter subtract the earlier year from the later year in order to obtain the change in each diversification score over time.

Tables 2 and 3 give results for all the diversification variables in a multivariate probit model, using five separate parliamentary divisions on the question of free trade. (The nature of each division is described in the appendix; simply put, divisions 1846a and 1846b are most critical, as the former is the first reading and the latter is the third reading of the repeal legislation.) M.P. party identification, predicted district trade preference (based on economic composition), and the effect of the 1832 Reform are included as control variables. I have explored the relationships between these variables and M.P. voting behavior on free trade in an earlier work;[48] consequently, the discussion here is limited to the diversification variables.

In brief, positive coefficients and significance for the level diversification variables (Tables 2 and 3) lend firm support to the hypothesis that M.P.'s representing constituencies with greater diversification were more likely to vote for free trade (that is, for repeal of the Corn Laws). The more comprehensive sample of income tax returns provides much higher levels of statistical significance than does the death duty sample. Nevertheless, obtaining as high as 95 percent significance (1846a) for such individual-level historical data as death duties is encouraging. Both samples exhibit the same trend: each begins with weak significance for diversification in 1834 and strengthens in significance in 1846, as one would expect if diversification were increasing during the 1830s and early 1840s.

Results for the difference variables (Tables 2 and 3) are far less straightforward than those for the level variables. Interpretation of these results draws upon earlier findings that established a classic urban-rural split between the voting behavior of borough and county M.P.'s, with the former significantly more likely to support free trade than the latter.[49] Results in Tables 2 and 3 are divided into subgroups of counties and boroughs in order to reveal the puzzling divergence between the two samples. That is, in the death duty sample, county M.P.'s do not appear

that factor returns—as capital flows—will shift from low- to high-yield sectors of the economy.

[48] Schonhardt-Bailey (fn. 1), chap. 3.

[49] Ibid.

to have responded to constituency diversification by voting for free trade, whereas their borough counterparts do seem to have responded. The results for the income tax returns reveal just the opposite trend: whereas county M.P.'s appear to have responded to constituency diversification with a vote for free trade, borough M.P.'s do not seem to have responded likewise. Before progressing further, it should be acknowledged that tests of statistical significance are unimpressive for both samples; interpretation must accordingly be speculative (especially for divisions 1842 and 1846c, where only two M.P.'s from county constituencies voted for free trade). Nevertheless, the relative internal consistency of both sets of results and the partial statistical significance lead me to venture a speculative interpretation.

Recalling that the death duty register sample measured solely the income from stocks and shares as sources of diversification, we may interpret the results in Table 2 to mean that the spread of stock market activity seemed to contribute to borough M.P.'s voting for repeal, but that this one aspect of diversification was not sufficient to dissuade county M.P.'s from their protectionist stance. However, if diversification is measured more broadly to include all trade and manufacture, as in the income tax returns sample, then the magnitude of diversification does appear to affect the voting behavior of county M.P.'s. The obvious question is, why did borough M.P.'s in the income tax returns sample not exhibit the same positive correlation? The answer simply put is that in both periods most boroughs were far more diversified into industry than were counties; thus, reduced but nevertheless still high diversification scores correlated negatively with free-trade votes on the part of borough M.P.'s. (Strong positive correlations between diversification and free-trade votes for the level variables lend support to this interpretation.) A similar result did not occur in the death duty register sample since (1) whole county averages were adopted for diversification indexes, rather than splitting county and borough constituencies, and (2) no doubt as a consequence of using whole county averages, initial scores were virtually all very low.

In sum, while borough M.P.'s appeared more responsive than county M.P.'s to diversification by their constituents into stocks and shares, county M.P.'s (and probably also borough M.P.'s) did seem to respond to more general diversification into trade and manufacture. Thus, for the difference variables, the type of diversification (specific, as in stocks, versus general, as in trade and manufacture) and the type of constituency represented (borough versus county) must be considered if the hypothesis of diversification is to be sustained.

TABLE 2

DEATH DUTY REGISTER SAMPLE: PROBIT RESULTS FOR LEVEL AND DIFFERENCE VARIABLES[a]

	All Constituencies					Counties			Boroughs		
Division	Level Coefficients[b]		(t-ratio)	Number of Cases	Difference Coefficients[b]		(t-ratio)	Number of Cases		(t-ratio)	Number of Cases
1834	Constant	−4.38	(−7.42)[f]	341	Constant	−10.39	(−1.25)	86	−4.21	(−6.85)[f]	218
	B1 DIVS23CO	0.24	(0.94)		B1 DIVSDIF2	−0.14	(−0.28)		−0.13	(−0.50)	
	B2 DIV2PID	0.81	(6.28)[f]		B2 DIV2PID	1.61	(0.88)		0.77	(5.59)[f]	
	B3 DISTPREF	0.33	(7.68)[f]		B3 DISTPREF	0.66	(3.78)[f]		0.20	(3.72)[f]	
	B4 REFCHNG	0.24	(2.92)[f]		B4 REFCHNG	0.81	(0.85)		0.30	(3.39)[f]	
1842	Constant	−8.66	(−4.16)[f]	326	Constant	−7.85	(−0.04)	95	−8.90	(−3.93)[f]	193
	B1 DIVS25CO	0.56	(1.26)		B1 DIVSDIF2	−0.72	(−0.72)		0.48	(1.26)	
	B2 DIV3PID	1.94	(3.92)[f]		B2 DIV3PID	2.52	(0.40)		1.93	(3.58)[f]	
	B3 DISTPREF	0.30	(3.95)[f]		B3 DISTPREF	0.25	(0.67)		0.25	(2.75)[f]	
	B4 REFCHNG	0.25	(1.87)[d]		B4 REFCHNG	−0.82	(−0.02)		0.29	(1.94)[d]	
1846a	Constant	−2.22	(−4.74)[f]	372	Constant	−2.85	(−0.86)	98	−3.11	(−4.78)[f]	234
	B1 DIVS25CO	0.67	(2.25)[e]		B1 DIVSDIF2	−0.13	(−0.28)		0.39	(1.36)[e]	
	B2 DIV4PID	1.13	(10.52)[f]		B2 DIV4PID	1.19	(5.06)[f]		1.05	(7.58)[f]	
	B3 DISTPREF	0.25	(4.70)[f]		B3 DISTPREF	0.34	(2.11)[e]		0.08	(1.14)	
	B4 REFCHNG	−0.08	(−0.72)		B4 REFCHNG	−0.17	(−0.21)		0.28	(1.63)[e]	
1846b	Constant	−2.21	(−4.66)[f]	367	Constant	−2.58	(−0.69)	95	−3.29	(−4.62)[f]	234
	B1 DIVS25CO	0.56	(1.88)[d]		B1 DIVSDIF2	−0.74	(−1.36)[e]		0.48	(1.65)[d]	
	B2 DIV6PID	1.15	(10.30)[f]		B2 DIV6PID	1.38	(4.82)[f]		1.05	(7.34)[f]	

1846c	Constant	−4.15	(−6.68)[f]	235	Constant	−7.18	(−0.03)	59	−4.03	(−6.65)[f]	154
	B1 DIVS25CO	0.55	(1.28)[c]		B1 DIVSDIF2	0.02	(0.02)		0.17	(0.48)	
	B2 DIV7PID	0.99	(6.84)[f]		B2 DIV7PID	2.68	(0.40)		0.88	(5.71)[f]	
	B3 DISTPREF	0.10	(1.71)[d]		B3 DISTPREF	0.15	(0.43)		0.04	(0.53)	
	B4 REFCHNG	0.10	(0.99)		B4 REFCHNG	−1.07	(−0.02)		0.16	(1.50)[c]	

[a] Probability of free-trade vote = A + B1 (diversification of constituents' income) + B2 (M.P. party ID) + B3 (predicted trade preference of M.P. constituency district) + B4 (effect of 1832 Reform on district).

[b] DIVS23CO gives the diversification scores for the 1830 death duty register sample, averaged for each county; DIVS25CO gives the diversification scores for the 1850 death duty register sample, averaged for each county; DIVSDIF2 is the difference between DIVS23CO and DIVS25CO. Variable descriptions are as follows: DISTPREF is predicted trade preference, based upon constituency district economic composition in 1831–32; REFCHNG is effect of the 1832 Reform Act on electoral constituencies; DIV2PID is party affiliation of M.P.'s, for 1834 division; DIV3PID is party affiliation of M.P.'s, for 1842 division; DIV4PID is party affiliation of M.P.'s, for 1846a division; DIV6PID is party affiliation of M.P.'s, for 1846b division; DIV7PID is party affiliation of M.P.'s, for 1846c division.

[c] Significant at the 80% confidence level, two-tailed test.

[d] Significant at the 90% confidence level, two-tailed test.

[e] Significant at the 95% confidence level, two-tailed test.

[f] Significant at the 99% confidence level, two-tailed test.

TABLE 3

INCOME TAX RETURNS SAMPLE: PROBIT RESULTS FOR LEVEL AND DIFFERENCE VARIABLES[a]

Division	All Constituencies: Level Coefficients[b]		(t-ratio)	Number of Cases	Difference Coefficients[b]	Counties	(t-ratio)	Number of Cases	Boroughs	t-ratio)	Number of Cases
1834	Constant	−5.58	(−4.60)[g]	214	Constant	−9.19	(−1.27)	110	−5.25	(−4.51)[g]	104
	B1 TXDIVS14	0.48	(1.81)[e]		B1 TXDVDIFF	2.63	(1.36)[d]		−3.45	(−1.78)[e]	
	B2 DIV2PID	0.81	(4.27)[g]		B2 DIV2PID	1.49	(0.95)		0.71	(3.53)[g]	
	B3 DISTPREF	0.24	(3.97)[g]		B3 DISTPREF	0.50	(3.47)[g]		0.15	(2.01)[f]	
	B4 REFCHNG	0.45	(1.72)[e]		B4 REFCHNG	0.58	(0.66)		0.67	(2.25)[f]	
1842[c]	Constant	−13.55	(−1.00)	227	Constant	−19.64	(−0.20)	126	−24.18	(−1.75)[e]	101
	B1 TXDIVS56	1.15	(2.45)[f]		B1 TXDVDIFF	−3.67	(−0.49)		−14.70	(−2.15)[f]	
	B2 DIV3PID	2.99	(0.88)		B2 DIV3PID	2.30	(0.39)		5.73	(1.67)[e]	
	B3 DISTPREF	0.16	(1.37)[d]		B3 DISTPREF	0.15	(0.40)		0.27	(1.80)[e]	
	B4 REFCHNG	0.24	(0.85)		B4 REFCHNG	2.41	(0.10)		0.47	(1.25)	
1846a	Constant	−5.24	(−3.68)[g]	239	Constant	−5.64	(−2.23)[f]	130	−5.07	(−2.65)[f]	109
	B1 TXDIVS56	1.24	(3.60)[g]		B1 TXDVDIFF	1.78	(1.10)		−4.20	(−1.53)[d]	
	B2 DIV4PID	1.21	(7.49)[g]		B2 DIV4PID	1.16	(5.51)[g]		1.36	(4.10)[g]	
	B3 DISTPREF	0.15	(2.10)[f]		B3 DISTPREF	0.39	(2.94)[g]		0.04	(0.47)	
	B4 REFCHNG	0.48	(1.40)[d]		B4 REFCHNG	0.50	(0.81)		0.76	(1.53)[d]	
1846b	Constant	−4.47	(−3.29)[g]	236	Constant	−5.85	(−2.17)[f]	127	−5.10	(−2.63)[f]	109
	B1 TXDIVS56	1.06	(3.16)[g]		B1 TXDVDIFF	0.51	(0.30)		−1.45	(−0.57)	
	B2 DIV6PID	1.18	(7.54)[g]		B2 DIV6PID	1.23	(5.52)[g]		1.36	(4.16)[g]	

1846c	Constant	−7.56	(−4.15)[g]	151	Constant	−3.49	(−0.02)	80	−5.83	(−3.90)[g]	71
	B1 TXDIVS56	1.96	(3.23)[g]		B1 TXDVDIFF	3.00	(0.32)		−1.38	(−0.36)	
	B2 DIV7PID	1.49	(4.74)[g]		B2 DIV7PID	3.27	(0.33)		1.46	(4.47)[g]	
	B3 DISTPREF	−0.30	(−2.18)[f]		B3 DISTPREF	0.01	(0.03)		−0.30	(−2.10)[f]	
	B4 REFCHNG	0.31	(0.88)		B4 REFCHNG	−2.69	(−0.05)		0.38	(1.06)	

[a] Probability of free-trade vote = A + B1 (diversification of constituents' income) + B2 (M.P. party ID) + B3 (predicted trade preference of M.P. constituency district) + B4 (effects of 1832 Reform on district).

[b] TXDIVS14 gives the borough and county aggregate diversification scores for the 1814 income tax returns sample; TXDIVS56 gives the borough and county aggregate diversification scores for the 1856 income tax returns sample; and TXDVDIFF is the difference between TXDIVS14 and TXDIVS56. Variable descriptions are as follows: DISTPREF is predicted trade preference, based upon constituency district economic composition in 1831–32; REFCHNG is effect of the 1832 Reform Act on electoral constituencies; DIV2PID is party affiliation of M.P.'s, for 1834 division; DIV3PID is party affiliation of M.P.'s, for 1842 division; DIV4PID is party affiliation of M.P.'s, for 1846a division; DIV6PID is party affiliation of M.P.'s, for 1846b division; DIV7PID is party affiliation of M.P.'s, for 1846c division

[c] In this division, DIV3PID (M.P. party affiliation) exhibited virtually no variation. In contrast to linear regression, this made obtaining statistical significance difficult in probit analysis.

[d] Significant at the 80% confidence level, two-tailed test.

[e] Significant at the 90% confidence level, two-tailed test.

[f] Significant at the 95% confidence level, two-tailed test.

[g] Significant at the 99% confidence level, two-tailed test.

CONCLUDING REMARKS

This paper focuses on the application of the specific factors model t
industrializing country, in this case Britain in the first half of the n
teenth century. I argue that the driving force behind the diversifica
of the asset portfolios of landowners was the rapidly emerging ca
market, evidenced in the 1830s by the spread of stock market acti
throughout Britain. I attempt to prove that the apparent difficult
reconciling conventional applications of the specific factors model
the decision of M.P.'s to repeal the Corn Laws can be overcome
switching from a model based on a measurement of the asset stock
voters to one that incorporates the corollary of industrialization, nam
the diversification of factor returns from land rents into higher y
industrial capital. This makes a model based on income flows rather t
on capital stocks a more meaningful tool.

The results at both the individual and the aggregate level reporte
this paper indicate that as landowners diversified into industry, the
litical benefits accruing to M.P.'s from maintaining protection in a
culture diminished. M.P.'s from districts that were more diversi
away from agriculture were more likely to vote in favor of free tr
than were those from less diversified districts. In turn, the diversificat
of districts was a measure of the degree to which individual voters w
acquiring a range of specific interests that were focused less on agri
tural rents. Thus, there was a lessening of the incentives for M.P.'
defend the Corn Laws and a decisive policy shift by Britain toward
eralization of trade.

Early industrializers such as Britain and the U.S. mobilized capita
smaller units, since the average scale of each business was smaller tha
was for later industrializers. In Britain, the combination of a large r
for direct investment (in the form of ownership of industrial product
by landowners) and the early development of securities markets enab
profit-maximizing landowners to switch directly into industrial inv
ment. It is the hypothesis of this paper that such direct ownership
industrial capital was an important factor in changing the trade pre
ences of landowners. Turning to cases of later industrialization (for
ample, Germany and Japan) in which larger sums of initial capital
vestment were required, the dominant actors responsible for managi
the flow of investment capital were banks. The requirement for lar
sums of investment capital created a role for banks as providers a
thereby limited the direct involvement of landowners in the diversifi
tion of their portfolios (though this does not provide an explanation
why in some countries banks rather than securities markets were t

major source of capital-raising activity). Landowners were able to invest in banks whose role was to mobilize the capital needed for larger units of production and that could pool the risks to an extent that could no longer be done by individuals investing directly. This paper has found a link between direct investment in industrial capital and a change in trade policy preferences; however, where the investment link is indirect, via a bank that pools risk by spreading its assets across economic sectors, it is more likely that the trade preferences of asset holders will be more weakly specified.

Appendix

Parliamentary Divisions

1826. Mr. W. Whitmore's motion for the consideration of free trade in corn (April 18, 1826): ayes = 81; noes = 250. (Division included only the minority M.P.'s and of these, only 65 could be identified by district.)

1834. Mr. Hume's motion for a committee on the Corn Laws (March 6, 1834). This was the first motion on the Corn Laws to be brought forth in the reformed Parliament: ayes = 155; noes = 312. (Two M.P.'s voting no could not be identified by district.)

1842. Mr. Villier's motion to abolish all the duties on corn (February 24, 1842): ayes = 90; noes = 393. (Two M.P.'s voting no could not be identified by district.)

1846(a). Mr. P. Mile's amendment to delay the House going into committee to consider the Corn laws by six months (thus forcing a vote on Peel's motion that the House immediately resolve itself into a committee of the whole, i.e., Peel's first reading of the Corn Law Repeal) (February 27, 1846): ayes = 337; noes = 240. (One M.P. voting aye could not be identified by district.)

1846(b). Mr. Peel's motion to repeal the Corn Laws, third reading (May 15, 1846): ayes = 327; noes = 229. (All M.P.'s identified by district.)

1846(c). Mr. Villier's amendment calling for total and immediate repeal of the Corn Laws (versus the gradualist nature of Peel's motion, with complete abolition of duties in 1849) (March 6, 1846). The House divided on the question that Peel's motion retain the 1849 date: ayes = 265; noes = 78. (Ayes coded as a vote for protection, and noes coded as a vote for free trade. Districts of two M.P.'s were questionable but were coded according to my best guess estimate.)

Review Articles

THE PEOPLE'S DEMOCRATIC REPUBLIC OF ETHIOPIA
Masking and Unmasking Tragedy

By FORREST D. COLBURN*

Christopher Clapham. *Transformation and Continuity in Revolutionary Ethiopia*. Cambridge: Cambridge University Press, 1988, 284 pp.
Dawit Wolde Giorgis. *Red Tears: War, Famine and Revolution in Ethiopia*. Trenton, N.J.: Red Sea Press, 1989, 375 pp.
John W. Harbeson. *The Ethiopian Transformation*. Boulder, Colo.: Westview Press, 1988, 239 pp.
Paul B. Henze. *Ethiopia: Crisis of a Marxist Economy: Analysis and Text of a Soviet Report*. Santa Monica, Calif.: Rand Corporation, 1989, 83 pp.
Edmond J. Keller. *Revolutionary Ethiopia: From Empire to People's Republic*. Bloomington: Indiana University Press, 1988, 307 pp.
John Markakis. *National and Class Conflict in the Horn of Africa*. Cambridge: Cambridge University Press, 1987, 314 pp.
Mekuria Bulcha. *Flight and Integration: Causes of Mass Exodus from Ethiopia and Problems of Integration in the Sudan*. Uppsala, Sweden: Scandinavian Institute of African Studies, 1988, 256 pp.
Mulatu Wubneh and Yohannis Abate. *Ethiopia: Transition and Development in the Horn of Africa*. Boulder, Colo.: Westview Press, 1988, 224 pp.
Roy Pateman. *Eritrea: Even the Stones Are Burning*. Trenton, N.J.: Red Sea Press, 1990, 239 pp.

ETHIOPIA—like China, Egypt, and Iran—is one of the oldest countries in the world. Its recorded history dawns in the millennium before Christ with the establishment of the Kingdom of Aksum, which was situated in the north of the country in what is now Tigray and Eritrea. Homer mentions Ethiopia in his chronicles. Indeed, the word *Ethiopia*, meaning land of the people with burnt faces, is, like the word *Egypt*, of Greek origin. At its zenith the Kingdom of Aksum was the most powerful state between the Roman Empire and Persia. Aksum's conversion to Christianity in the fourth century led to enduring ties with the Coptic church of Egypt and sporadic contact with Byzantium. Early

*Helpful comments on an earlier draft were offered by Henry Bienen, Milton Esman, Jeffrey Herbst, James McAdams, and Norman Uphoff.

European travelers were fascinated by Ethiopia and publicized the *Kibre Negest*, the epic that describes the founding of the Ethiopian nation, the glory of its monarchs, and their descent from King David of Israel through the union of Queen Sheba and David's son Solomon. Their son was Menelik I, king of Ethiopia. Emperor Haile Selassie I of Ethiopia, the Conquering Lion of the Tribe of Judah, King of Kings, Elect of God, traced his lineage to Menelik I, and so, too, to King David.

On the eve of the Ethiopian New Year in September 1974, Emperor Haile Selassie, nearing eighty-two, was bundled into a Volkswagen by a group of military officers and driven away from his palace to imprisonment. Since then Ethiopia—a country the size of France and Spain combined and the third most populous state on the African continent—has experienced a profound social revolution. The results have been horrifying. Probably nowhere else in the world is there such suffering. The government is guilty of gross human rights violations, including large-scale murder. The multinational polity is being wrecked by ever-increasing ethnic warfare. The carnage is overwhelming. The World Bank holds that per capita income has fallen steadily and that Ethiopia, always poor, is now the poorest country in the world. In 1984–85 a million Ethiopians starved to death, and only massive foreign assistance prevented the death of another eight million. Famine continues. Although only an estimated 12 percent of the population has ready access to roads, there are over two million Ethiopian refugees. What has happened? And why?

Recent books on Ethiopia answer the first question but stumble on the second. Ethiopia as a country and a polity presents numerous difficulties to scholars seeking to describe and explain what has befallen the nation-state. The central outline of events is transparent in recent scholarship. But the books that should be the most insightful, because of the theoretical ambitions of the authors, often disappoint because they obfuscate politics with abstract discussions of the "state." Discussions too quickly leap to a higher level of generality than warranted. The starkness of Ethiopia's politics demands analytical constructs that cling closer to political discourse and practice.

Mulatu and Yohannis[1] offer a descriptive account of Ethiopia that includes a succinct narration of how the military officers who ousted the emperor consolidated power and committed Ethiopia to Marxism-Leninism. Economy and society were reorganized according to the dictates of "scientific socialism." Slowly, the new regime institutionalized

[1] Ethiopians are usually referred to only by their first name.

itself: a loose committee of military officers under the leadership of Major Mengistu Haile-Mariam evolved first into the Provisional Military Government of Socialist Ethiopia and in turn into the People's Democratic Republic of Ethiopia. Comrade Mengistu's title evolved correspondingly: general secretary of the Central Committee of the Workers' Party of Ethiopia, president of the People's Democratic Republic of Ethiopia, and commander in chief of the Revolutionary Armed Forces. In the most prosaic language, Mulatu and Yohannis also describe the two most pronounced outcomes of government initiative: ethnic rebellion and economic decline.

More theoretical works, such as those by Clapham, Keller, and Harbeson, strive to explain what has happened in Ethiopia. These three scholars openly share a theoretical reference: Theda Skocpol's comparative study of revolution in France, Russia, and China.[2] All speak approvingly of this study but suggestively appear to move away from Skocpol's structuralist interpretation of revolutions, which has its roots in Marxist analysis of class and capital. Ethiopia's leaders have drawn on the same Marxist fountainhead but more openly and with less nuance. Clapham is the bluntest about rejecting Marxist-inspired theories of revolution, even for regimes professing Marxism. He suggests an irony: "In an age in which revolution is almost invariably associated with Marxism (quite regardless of the inadequacy of Marxist precepts in explaining its occurrence)" (p. 6), there is no "necessary connection between revolution and any particular form of economic structure" (p. 2).

Clapham argues that it is political ineffectiveness, rather than economic exploitation, which leads to revolution. But what makes a revolution is not the destruction of the old order but the construction of the new one—of a new *political* order. Clapham holds that the primacy of the political is equally central to postrevolutionary institutionalization: "It is organized political power, in the hands of the new rulers of the state, that has to be used . . . to bring about a deliberate transformation of economy and society which would not take place on its own" (p. 7).

Clapham finds Skocpol's approach most compelling in its emphasis on the centrality of the control of state power in the early phase of the revolution and of state building in the latter phase. In this latter phase, Clapham takes revolutionaries' Leninism seriously: "The Leninist principle of democratic centralism appears to provide an extraordinarily effective mechanism for combining tightly centralized elite control with the cooptation of the able and ambitious, and at least a token level of mass partic-

[2] Skocpol, *States and Social Revolutions* (Cambridge: Cambridge University Press, 1979).

ipation and democratic accountability" (p. 9). In contrast, Clapham dismisses ideology: "Ideology is generally (as Marx quite correctly identified it) a mere rationalization of the interests of its proponents" (pp. 9–10). Frankly, he says, "I remain unpersuaded of the centrality of ideology to revolutionary reconstruction" (p. 10).

Clapham has two other bones to pick. There is, he says, a misguided tendency to concentrate on the early, violent, and dramatic phase of the revolution, to the neglect of the period of institutionalization that follows. A second common weakness in the study of revolution is almost the converse of the first: a neglect, not only of what follows the first and heroic phase, but of what precedes it. "It is then all too easy to ascribe any particular manifestation of post-revolutionary attitudes or practices to the effects of revolution in itself, whether these be regarded as good or bad, while failing entirely to recognize the extent to which they duplicate pre-revolutionary equivalents" (p. 13).

Clapham's narration of the Ethiopian Revolution is richly detailed. But his analysis—and conclusions—are bound by an old-fashioned structural functionalism that often appears wooden. The following sentences are illustrative:

> It fell, not because it was badly damaged, but because it was by nature incapable of handling the increasing demands which were placed on it. (p. 32)

> In a society founded on a Hobbesian conception of the centrality of power relations to the maintenance of public order, the removal of a once-dominant authority figure always carried the threat of anarchy. (p. 43)

> In Addis Ababa, the failure of imperial authority unleashed an upsurge of demands which could not simply be suppressed, but for which some outlet had to be found. (p. 43)

All of these statements may, in some sense, be true, but they give little play to the drama of politics, of the centrality of choice. Slighted, if not ignored, are the influences of personality, political culture, ideas, leadership, organization, and institutions. Clapham comes close to succumbing to the same determinism for which he so persuasively chastises Marxists, only here the villain is not economic structures but the imperatives of government.

Not surprisingly, Clapham's conclusions are timid. There is no indictment. The first sentence of his conclusion reads, "Since 1974, and especially since the stabilization of the revolutionary regime in 1977–78, Ethiopians have found themselves part of a highly effective, disciplined and centralized state system" (p. 241). Well, yes. Only at the very end of

the book does he suggest that the coincidence in Ethiopia of the most powerful African state and the poorest African economy raises questions about the adequacy of the state as a motor of development. Continuing the metaphor (and the reification), he refers to the state as a machine:

> It is run for the most part efficiently, and often indeed with dedication. But the construction and operation of such a machine requires the subordination of all else to the needs of the machine itself, and especially its demands for control, in a way which ... often impedes production and fosters resistance. (p. 243)

The state—and those at its helm—are masked. No one is responsible for anything.

Harbeson, too, centers his analysis of Ethiopia on the state. He shares many of Clapham's theoretical precepts, including something of Clapham's structural functionalism. But the two dispute the strength of the postrevolutionary state: Clapham judges it strong; Harbeson claims that it is weak, even more so than the imperial regime. Accordingly, Harbeson argues that Ethiopia mirrors the course of conflict between the military regime and its adversaries. The military regime's initiatives have provoked multifaceted crises over the transformation, objectives, and even existence of a postimperial state.

Clapham and Harbeson disagree about the state largely because they define it differently: Clapham comes close to equating it with government; Harbeson follows Max Weber, who suggested that governmental institutions are not synonymous with the state but are derived instead from an underlying, generally recognized basis for political community (be it only shared territory).[3] Harbeson gives curt treatment to many political variables, but his broad definition of the state enables him to highlight conflict and, in particular, conflict for power. And he posits a curious, if murky, nexus between a strong government and a weak state.

Keller goes to greater lengths to disaggregate the state than either Clapham or Harbeson. Hence, he engages in less reification. Political actors are perceived as having choices and being influenced, not just by inert political economic structures, but also by ideas. Indeed, in seeking to explain the Ethiopian Revolution, Keller harks back to Crane Brinton's suggestion of the importance of "the allegiance of intellectuals."[4] Before the monarchy could be deposed, there had to be an alternative conception of how Ethiopia could be ruled. That alternative came from Ethiopian university students who had been radicalized in the

[3] Weber, *The Theory of Social and Economic Organizations*, trans. Talcott Parsons (Glencoe, Ill.: Free Press, 1947), 156.

[4] Brinton, *The Anatomy of Revolution* (1938; reprint, New York: Vintage, 1965), 39.

60s. Once the emperor was overthrown, there was "the challenge of :roducing a new ideology, a new social myth, that would at once pro- le the basis for its legitimacy among the general population and serve a guide for policy and political action" (p. 191). Students and the mil- ry engaged in a furious struggle for command of the state. Inevitably, e military won, but the students provided the new "social myth"— arxism-Leninism.

Why did the military adopt the students' Marxism-Leninism? Was it cause many of the younger military officers who staged the coup were io radicalized at Haile Selassie I University or elsewhere? Was it be- use the military officers genuinely believed students' rhetoric? Was it cause they had no ideas of their own? Was it because it was a clever ategy for undercutting the appeal of the students, the only immediate 'al to military leadership? Or was it because Marxism-Leninism justi- s a ruling oligarchy and a statist development strategy?

These questions elude definitive answer. But it seems unnecessarily oss to dismiss Marxism-Leninism, as Clapham does, as only providing nography (p. 97). At the least, Marxism-Leninism explains much of e *content* of the Ethiopian Revolution. Many institutions and policies e inspired by Marxism-Leninism or slavishly copied from regimes pro- ising Marxism-Leninism. Indeed, in Ethiopia some of the regime's ost consequential policies came directly from the inexperienced but ilitant students: the literacy campaign, the agrarian reform, and villag- ition. The institutions and policies shaped by Marxism-Leninism are t only important in their own right, but they also often have a decisive pact on the welfare, life chances, and political responses of different cial groups. And the embrace of Marxism-Leninism by the country's idership, whatever the motivation, has an impact on its ability to wield wer. As the cagey Haile Selassie himself knew, certain ideologies, or cial myths, facilitate one kind of rule as opposed to another.

If the ideas that inspire the "state" are to be taken seriously, so, too, ose who actually make the decisions on which the state stands or falls ould be taken seriously. Who are the leaders and how do they rule? ow do powerful individuals guide and mediate institutional politics? iis imperative is all the more compelling when, as in Ethiopia, a single lividual all but publicly exclaims: *l'état, c'est moi*! Here Clapham, arbeson, and Keller alike do not satisfy. Yes, all three describe Com- de Mengistu as ruthless, and all do trace his ascendancy to power. But ere is not a sense for how Ethiopia is different for his singular presence. ie difference between Mengistu and the government is not clarified. iis and similar omissions are troubling.

A very different book, one that is most compelling, is Dawit's memoir *Red Tears*. At the time of the emperor's ousting, Dawit Wolde Giorgi was a former military officer serving as deputy chief of the Relief and Rehabilitation Commission. During the shift in Ethiopia's alignment from the United States to the Soviet Union, he served as deputy foreign minister. Subsequently, he became a member of the Central Committee of the Workers Party of Ethiopia and served as its principal political representative in Eritrea. He headed famine relief during the great famine of 1983–85. He left Ethiopia in 1986.

Given Dawit's background, one would expect a good story. And he does offer an engaging story. Eritreans have a decidedly less charitable interpretation of his tenure in Eritrea than he offers, but otherwise his tale is persuasive. Dawit covers the same ground tackled by his academic counterparts: the demise of imperial rule, the emergence and institutionalization of the new regime, ethnic conflict, and—above all—the famine. Predictably, his work is rich in details that only an insider could know:

> The hall where Mengistu holds large conferences is ironically the same hall where Emperor Haile Selassie used to give his big banquets—elaborately redecorated with Marxist icons, of course. Mengistu once confided to me that he enjoyed chairing meetings in this hall because he was able to sit right above the basement where all the former aristocrats whom he despised were imprisoned. As usual, Mengistu sat on an elevated, red velvet seat, not far from the throne of the former Emperor, now covered with a white sheet. I had always wondered why the throne had not been taken to the museum with all the rest of the imperial trappings; why it was still here draped in this sheet, as if its owner were merely off on a long holiday and might return at any moment. (p. 126)

But what is so striking about Dawit's book, for the academy at least, is the descriptive and explanatory success he achieves by eschewing the broad concept of the state.

As he links his own life to the unfolding of the Ethiopian Revolution, Dawit routinely explains how key decisions were made and with what effect. And in focusing on choices, he effortlessly shifts the "unit of analysis" as the need arises; he moves, for example, from a single individual—usually Mengistu—to bureaucratic infighting to broad social conflict. His inductive approach yields a revealing, complex portrayal of the Ethiopian Revolution. Three of the highlights are (1) the importance of ideology and the extent to which it was provided by the university community, (2) the dominance of Mengistu, and (3) the unnecessary suffering of Ethiopians, ranging from the dozens of Mengistu's fellow military

officers whom he executed to the unnamed masses of peasants who are victims of ill-conceived, counterproductive agrarian policies.

In comparison, the fixation on a single unit of analysis—the state—by Clapham, Harbeson, and, to a lesser extent, Keller, obscures these important features of the Ethiopian Revolution. Blame can perhaps be attributed to an unwarranted emulation of the theoretical precepts of heralded comparative studies, exemplified by the work of Crane Brinton, Barrington Moore, and Theda Skocpol.[5] Their books certainly deserve respect. It is fruitful to trace broad historical changes. But the theoretical parameters necessary for that kind of sweeping endeavor are probably ill suited to the analysis of a single, contemporary revolution. Speculating on the extent to which an incipient postrevolutionary regime is fulfilling the traditional responsibilities of the state is less useful than dissecting what kind of state it is in the first place. In short, focusing on the government is probably more useful than focusing on the state.

The primary reason for examining a revolutionary regime with more specificity and more attention to the conduct (as opposed to the outcome) of politics is to understand it more accurately. The nation-state can still be the subject of inquiry; there is no need to retreat into micro studies. And there is concurrently no need to retreat from theorizing. There are many puzzles worthy of explanation. Invoking the state—exploring its strength and autonomy—may prove necessary. But care is needed to correct what Sartori once called "conceptual stretching," by providing specificity to the logical and empirical properties of the state.[6] And care must be taken to guard against state centrism.

A second reason to disaggregate the "state" is to identify who bears responsibility for the tragedy, or tragedies, that Ethiopians have suffered. In a macrohistoric, comparative study of revolution, it is perhaps permissible to gloss over such cases as the Soviet famine of 1932–34 that killed five million or the Chinese famine of 1958–60 (the "Great Leap Forward") that killed ten million. But in contemporary Ethiopia—or Cambodia—where those responsible for famine are still potent political actors, it does not seem reasonable to speak so abstractly of governors.[7] Scholars may not have, or want the responsibility of judging right from wrong and of fingering the perpetrators of acts of gross immorality. Yet at the least, their work should make it clear who did what and with what

[5] Ibid.; Moore, *Social Origins of Dictatorship and Democracy* (Boston: Beacon Press, 1966); Skocpol (fn. 2).

[6] Giovanni Sartori, "Concept Misinformation in Comparative Politics," *American Political Science Review* 64 (December 1970), 1033–53, at 1034.

[7] For a comprehensive study of the Kampuchean regime, see Karl Jackson, ed., *Cambodia 1975–1978* (Princeton: Princeton University Press, 1989).

consequences for human welfare and life. Likewise, ideas, political passions, and ideologies that contribute to misery should be identified.

ETHNIC REBELLION

Intertwined with the Ethiopian Revolution are dissident nationalist movements, or national liberation fronts, as they call themselves. The strongest such movement, the Eritrean, dates back to 1960 with the founding of the Eritrean Liberation Front (ELF) by Muslim Eritreans educated—and politicized—in Nasser's Egypt. Haile Selassie's inability to solve the "Eritrean problem" led to dissatisfaction with the emperor within the ranks of Ethiopia's military. So Eritrean nationalism, and the war accompanying it, was a contributing cause of the revolution. But since the emperor was deposed, Eritrean nationalism has only swelled and has been matched by other dissident nationalist movements, most prominently in Tigray.[8]

The burden of ethnic strife is staggering. Impoverished Ethiopia is home to the world's longest running civil war—thirty years. Ethiopia marshals the largest military in Africa, and Eritrea has the eleventh largest army (Pateman, 121). Casualties from single battles have on occasion run into the tens of thousands. Much of northern Ethiopia is routinely either contested or controlled by the Eritrean or Tigrean liberation fronts. And there are sporadic efforts to heighten the national passions of other ethnic groups, including the Somali, the Afar, and the Oromo.

Clapham, Harbeson, and Keller discuss Ethiopia's ethnic strife. Suggestively, though, their efforts are limited to description. The three authors are versed in and committed to comparative analysis and social theory. But their interpretation of the role of comparative analysis and social theory in a single-country study is in overarching questions of the state. Unfortunately, they have not availed themselves of even the most celebrated comparative studies of ethnicity.[9] Had they set Ethiopia's ethnic strife within the framework of the literature on ethnicity, there would have been more depth to their treatment of this most vexatious problem. And they would likely have been prodded to delineate the

[8] Many names in Ethiopia do not have consistent spellings. Tigray is also written Tegray, Tegrai, and Tigre.

[9] The kind of work that could have been fruitfully consulted include Donald Horowitz, *Ethnic Groups in Conflict* (Berkeley: University of California Press, 1985); Walker Connor, *The Nationalist Question in Marxist-Leninist Theory and Strategy* (Princeton: Princeton University Press, 1984); Milton Esman, "The Management of Communal Conflict," *Public Policy* 21 (Winter 1973), 49–76; Eric Nordlinger, *Conflict Regulation in Divided Societies*, Occasional Paper no. 29 (Cambridge: Center for International Affairs, Harvard University, 1972).

causes of ethnic rebellion and to consider the extent to which Ethiopia's strife is typical or atypical of other, similarly situated regimes.

The hopelessly inconclusive arguments for Ethiopia's "territorial integrity" and for Eritrean "self-determination" are spelled out, respectively, by Dawit and Pateman. Despite Dawit's disdain for the Ethiopian government, he supports its efforts to combat secessionist movements, arguing, for example, that Eritrea is historically and culturally an integral part of Ethiopia. He suggests that a sense of unity transcends differences of nationality. Indeed, Eritrea itself is composed of nine nationalities and is equally divided between Christians and Muslims. Dawit quotes foreign historians who maintain that by virtue of a shared national heritage Ethiopia managed to survive the forces of internal disintegration and remain the only part of Africa not colonized by Europeans.

The Italians did try to conquer Ethiopia. And although their advances were halted, they occupied the northern coast—Eritrea—for fifty years. Dawit decries Italian colonialism, which included a decree that "natives must take their shoes off when they go into government offices" (p. 75). Italian rule ended with British occupation in 1941. When the British left, they dismantled and took with them port facilities, railway rolling stock, factories, and equipment. In return they engineered a UN General Assembly recommendation that Eritrea should "constitute an autonomous unit federated with Ethiopia under the sovereignty of the Ethiopian crown" (p. 79). Haile Selassie tried to yank the province back into his feudal domain.

The first sentence of Pateman's book reveals his perspective:

> Thinking back about the reasons for my interest in Eritrea, it seems that while a very young man of fifteen, I must first have read articles sympathetic to the Eritrean cause in British leftist journals written before the USSR decided to support Ethiopia's annexation of the country. (p. vii)

What follows is a well-researched, well-reasoned case for Eritrean independence. Arguing that "the struggle for independence also involves a struggle over history and the interpretation of the past," Pateman musters data showing that the histories of Eritrea and Ethiopia have often diverged (p. viii). He contends that Ethiopia has always been, and continues to be, disproportionately dominated by one ethnic group—the Amhara.

Pateman recounts the history of the ELF and of how it violently came to be replaced by the Eritrean People's Liberation Front (EPLF). (Ethnic parochialism was the first contradiction to emerge in the "national" liberation movement.) Much of the book is a grisly account of the chronol-

ogy of battles. In his conclusion, Pateman addresses one of the persistent international objections to Eritrean independence: small states are not viable. Here he is most persuasive:

> [Since UN mandated federation] ... over fifty countries with a smaller population and many with much fewer resources than Eritrea have become independent. In Africa, there are eighteen member states of the OAU with smaller populations than Eritrea; moreover some twenty micro states have been created in the Pacific and Caribbean. (pp. 211–12)

Small but viable states are held to be better than sprawling empires poorly governed and beset by domestic strife.

Markakis's study of nationalism in the Horn of Africa offers a more dispassionate analysis of the origins and evolution of dissident nationalism. Geographically and temporally his focus is broad. Following a profile of the region, successive chapters explore specific political movements "whose goal is to change the structure of the existing states in the Horn of Africa, or to establish states of their own" (p. xvii). Rich opportunities are afforded for comparative analysis.

Although Markakis shows the material inducements for dissident nationalism—most prominently, discrimination in government employment—he also points out that the leadership for both the Eritrean and Tigrean movements came almost entirely from the student population and was highly ideological:

> Eritreans in the university at Addis Ababa distinguished themselves by their political activism, and their immersion in the radical ideological currents that were beginning to surge in Ethiopian student circles in the mid-1960s. ... Many even among the older generation of educated Eritreans were to follow the example of the students who rallied to the nationalist movement almost as a body in the late 1960s. (p. 119)

Much discussed by students were Marx's views on the class origins of ethnic grievances and Lenin's views on the right of self-determination. Dawit reports that most Ethiopian students supported the Eritrean liberation movement because it was the only organized movement against Haile Selassie. According to Dawit, most reasoned that once Haile Selassie and the feudal system were overthrown, the national question and the question of self-determination would be solved by the eradication of class differences and the provision of equality and autonomy to all nationalities.

To students in the 1960s Ethiopia must have looked like the Russia Lenin denounced as "the tsar's prison of nations." But Lenin's writing on the nationalities question and the right of succession are ambiguous,

not contradictory. Ugly disputes broke out over the correct interpre-
tion, and discord mounted over the course of the Ethiopian Revolution.
iling to get what they wanted to Addis Ababa, students retreated to
eir native provinces and fanned dissident nationalism.

This interpretation of dissident nationalism, which emphasizes stu-
nts and their ideas, is supported by Markakis's account of the founding
the second most powerful liberation movement, the Tigrean:

> Tigrai youths in the university in Addis Ababa were active in the student movement and were well represented in radical circles. . . . [T]he Tigrai radicals became involved in a running debate with their compatriots . . . on the relative importance of class and national contradictions. . . . After the seizure of power [by the military] they produced a pamphlet . . . with the ominous title "E tek" (To Arms) . . . and began forming an organization . . . to struggle for self-determination. The die had been cast precipitately it would appear, for the Tigrai radicals challenged not only the newly established military regime . . . but they also broke ranks with their Ethiopian radical comrades. . . . In emulation of the Eritreans . . . first base was established on a mountain. . . . The emerging guerrilla force . . . [consisted of] a hundred youths, most of them former students. (pp. 252–53)

nce again, ideology seems to provide more than iconography. And the
eology that resonated was Marxism-Leninism. Ironically, it emerged
t from the university community's contact with the Soviet Union or
ith other socialist countries, but from ties to the U.S. and, to a lesser
gree, to Western Europe.[10]

What remains unclear from recent scholarship is the extent to which
e students' success can be traced to ethnic grievances as opposed to
neral dissatisfaction with Mengistu's regime. There must be some-
ing deficient or provocative in the way those who have governed this
ultinational system succeeded in fueling these rebellions. Unfortu-
ıtely, no scholar provides much illumination on this key, albeit difficult,
estion.

Dawit is exceptional in this regard in that, for him, the resolution of
hnic rebellion depends, not on the Ethiopian "state," but on the caprice
a single individual—Mengistu. In his concluding discussion of the
ssident nationalist movements, Dawit notes:

> From my observation of the capacity of these movements, their aspirations . . . the worst is yet to come and could lead to the breakup of the nation. Anarchy is descending over Ethiopia. In meeting after meeting, Mengistu has been told both politely and emotionally that we must look for alter-

[10] See Forrest Colburn, "The Tragedy of Ethiopia's Intellectuals," *Antioch Review* 47 pring 1989), 133–45.

natives to the military solution. He refuses and always predicts tha
situation will get better soon. It always get worse. (p. 119)

ECONOMIC DECLINE

Even before the military officers who ousted the emperor com
themselves to scientific socialism, they began seizing control of the
manding heights" of the Ethiopian economy. Industrial, financi
commercial enterprises were nationalized. And the state assume
control over foreign trade. In addition to nationalizing major
based firms, five thousand commercial farms were nationalize
though they represented only a small percentage of land under c
tion, commercial farms monopolized the production of such imp
crops as sugar and cotton. Finally, the government conferred or
broad authority to regulate what remained of the private sector
begin laying the foundations for central planning.

Arguably more important, however, have been government init
toward peasant agriculture. Ethiopia is not only a decidedly ag
economy but one dominated by peasants. Agriculture provides 8
cent of employment and 85 percent of exports. Peasants farm ab
percent of cultivated land. They produce 90–94 percent of the c
pulses, and oilseeds, and about 98 percent of the coffee, the most i
tant export.[11] Given the incipient level of modernization in Et
progress depends overwhelmingly on the health of agriculture
versely, any failure in the sector results in hardship because the b
the population lives so close to subsistence.

In early 1975, the new regime promulgated land reform: all rura
was nationalized without compensation; tenancy was abolished; th
ing of wage labor on private farms was forbidden; all commercial
were to remain under state control; and all peasants were to have
sessing right" to a plot of land not exceeding ten hectares (ro
twenty-five acres).[12] The reform was welcome in the southern p
Ethiopia, where tenancy was common and rural elites were exploi
But there was some resistance in northern provinces, where comr
ownership was typical and large holdings and tenancy were excep
The regime promptly organized peasants into peasant associatio

[11] Figures are from Steven Franzel et al., "Grain Marketing Regulations: Impact o
ant Production in Ethiopia," *Food Policy* 14 (November 1989), 347–58, at 348.

[12] The Provisional Office for Mass Organizational Affairs; Agitation, Propaganda a
ucation Committee, *Basic Documents of the Ethiopian Revolution* (Addis Ababa: Pro
Office for Mass Organizational Affairs; Agitation, Propaganda and Education Com
1977), 18–47.

ght of the poverty of Ethiopia and the inexperience of the new govern-
ient, the successful organization of peasants in every corner of the coun-
y outside those areas held by rebels was an extraordinary accomplish-
ient, even if it was done with the threat of coercion to those expressing
luctance. Initially, it appeared as if the peasant organizations would
rovide an institutional basis for rural development and self-help activi-
es. They have been used mostly as an instrument for government poli-
es motivated by a quixotic combination of ideological impulses and a
ragmatic need to extract resources.

All macroeconomic indicators suggest that the Ethiopian economy has
eteriorated in the postrevolutionary epoch. The poor economic perfor-
iance can be explained in part by severe resource constraints, aggra-
ated by secessionist movements, conflict with Somalia, drought, and
orld economic trends. Yet the depth, persistence, and uniformity of
onomic difficulties suggests that government policies are at fault, too.
roblems have emerged even on state farms where the government has
onsiderable autonomy and is well endowed with resources. An Ethio-
ian study reports:

> During a working visit ... Chairman Mengistu ... noted with heavy concern the shortcomings of the state farms and how the performance and operations of the state agro-industrial activities were far from desirable and at times catastrophic.[13]

ince 1974 annual growth in agricultural production has only been about
percent. Population growth has been approximately three times as
igh, resulting in declining per capita production.

Surveys of peasants have revealed that the production of foodstuffs
as been thwarted by forced requisitions of harvests at low prices and by
nwelcome efforts to promote collective farming.[14] And the nationali-
ation of land means that no household has any incentive to improve the
ind in ways that would safeguard, or increase, its productivity. Peasant
ssociations are unable to influence government policy, leaving peasants
ith no alternative except flight or foot dragging. Proverbial expressions
f resignation abound: "The son of the Blue Nile is thirsty"; "Whether
iere are one or two lactating cows, my cup of milk is the same, says the
at"; "May our great God help us."

Paul Henze, a frequent commentator on Ethiopia, presents in pub-

[13] Johannes Kinfu, "Towards Understanding the Public Corporation, and/or Public En-
rprise, State Enterprise in a Socialist Industrial Transformation and to Provide Proper
ccounting for It" (Mimeo, Addis Ababa, 1980), 25.

[14] See, for example, Dessalegn Rahmato, *Agrarian Reform in Ethiopia* (Trenton, N.J.: Red
ea Press, 1985); and Franzel (fn. 11).

lished form a lengthy Soviet report on the Ethiopian economy. The report was authored by advisers from the Soviet State Planning C mission (GOSPLAN). As with all correspondence between the Soviet U and Ethiopia, it was written in English. As Henze points out in hi troduction, the report was prepared before Gorbachev's reforms implemented; indeed, Gorbachev had been general secretary foı than six months at the time the report was completed. Yet the r criticizes the Ethiopian regime's reliance on state control of the econ disinterest in the private sector, neglect of prices, and attempted cc tivization of the peasantry. The hallmarks of Marxist-Leninist econo were dismissed as being counterproductive by the official representa of the most advanced—and experienced—socialist state.

Mekuria examines the fate of those who have fled Ethiopia, usual walking to neighboring Sudan or Somalia, two countries with enorı problems of their own. He claims that Ethiopian refugees constitute of all refugees in Africa. The tale of Ethiopian refugees is sad. Flig an option only for those who live close to the borders, and it is exc ingly risky. For those who reach the border, other difficulties pr themselves. In one survey of settlements in Sudan, 30 percent of the dren died during their first year in the country (p. 141). Integration other societies is elusive.

More commonly, desperate peasants migrate from the heart of country to the cities. Ethiopia's noted scholar Dessalegn Rahmato w passionately of these migrations:

> The mass migration of peasants to the urban areas carries . . . a variety meanings and messages. To begin with, it is a form of collective articul tion of the demand for the consecration of the right to life, a right whi requires the satisfaction of the most elementary needs of individuals a communities, namely food. Secondly, it is a form of silent protest. T migration of thousands of peasants from the vast recesses of the countr side where they are "invisible" to the road side towns there they are high visible contains a clear but unexpressed message: state and society have political or moral right to write us off, or to ignore our plight![15]

Given the secrecy of the Ethiopian government, the size and dive of the country, and the paucity of reliable data even in the desks of ra ing bureaucrats, the books reviewed here do a commendable job of lineating the economic woes of postrevolutionary Ethiopia. But ag the lack of comparative or theoretical analysis of specific characteri

[15] Dessalegn Rahmato, *Famine and Survival Strategies* (Addis Ababa: Institute of Dev ment Research, 1987), 188–89.

of the polity, such as economic mismanagement, is disappointing.[16] Abstract discussions of "state strength" or "state softness" are a poor substitute.

Conclusion

Studies of individual countries remain the most common work of scholars in the field of comparative politics. The motivation and rationale for the study of particular countries are clear and enduring. But prestige and sometimes, unfortunately, even legitimacy within the ranks of scholars is accorded to those who engage in cross-national studies. Even scholars who write about China, a country of over a billion inhabitants, are sometimes asked to justify "the larger significance of their work." To enrich their studies, satisfy their own curiosity, and disarm critics, scholars of individual countries routinely engage in comparative analysis. They do so with varying levels of commitment and skill. But there does seem to be a tendency, exemplified by recent scholarship on Ethiopia, to follow the dictates of academic fashion and pitch comparative inquiries to the highest level of generality—the state and the state system.

Grand statements about the state may well have utility for comparative macrohistory and for overarching societal theory. They are, however, entirely unsuitable for a single-country study. These must first and foremost present solid information about what really happened in the specific context, identifying the principal actors—individuals and groups—and analyzing and attempting to account for their behavior. Comparative analysis and social theory can be helpful, even necessary, for the task. But both are most profitably employed to help illuminate specific phenomena and processes, such as the conduct of government, communal politics, and economic management. The available "literature" encompasses far more than billowing discussions of the state.

Less abstraction is likely to result in more penetrating studies of individual polities. It is also likely to raise, on occasion, the hitherto neglected question of moral responsibility, as "sweeping historical forces" are replaced by choices made by individuals or groups. Contemporary African states are commonly said to be more or less, mostly more, misgoverned by incompetent and predatory rulers—the consequence of colonialism or neocolonialism, lack of indigenous institutions, insufficient experience, and cultural disabilities. According to this logic, Ethiopia is just another

[16] An example of the kind of comparative analysis that could be profitably consulted is David Morawetz, "Economic Lessons from Some Small Socialist Developing Countries," *World Development* 8 (May–June 1980), 337–69.

victim of these powerful, impersonal forces. Such explanations are suspect. The scholarship on Ethiopia reviewed here, despite its shortcomings, suggests an unrelieved chronicle of misgovernment that ought to be charged to those who have perpetuated it. Neither the brutal and continuing violations of human rights, nor the ethnic rebellions, nor the collapse of the economy were foreordained or historically inevitable. They were caused and certainly aggravated by Ethiopia's ruling elites, including the tragic conjunction of a military clique imposing Stalinist patterns of economic management. If students of politics are to regard political actors as morally responsible individuals, rather than products of impersonal historical forces, then the quality of their stewardship is indeed a matter that scholars should be willing to observe, analyze, and, yes, judge. Elites may be constrained by their circumstances, but they retain large degrees of freedom. And for the exercise and consequences of that freedom they should be held morally responsible.

NUCLEAR STRATEGY AND NUCLEAR WINTER

By THEODORE RUETER AND THOMAS KALIL

)avid E. Fisher. *Fire and Ice: The Greenhouse Effect, Ozone Depletion and Nuclear Winter*. New York: Harper and Row, 1990, 232 pp.

,ester Grinspoon, ed. *The Long Darkness: Psychological and Moral Perspectives on Nuclear Winter*. New Haven: Yale University Press, 1986, 213 pp.

\llen Lynch. *Political and Military Implications of the "Nuclear Winter" Theory*. New York: Institute for East-West Security Studies, 1987, 49 pp.

:arl Sagan and Richard Turco. *A Path Where No Man Thought: Nuclear Winter and the End of the Arms Race*. New York: Random House, 1990, 499 pp.

'eter C. Sederberg, ed. *Nuclear Winter, Deterrence, and the Prevention of Nuclear War*. New York: Praeger, 1986, 202 pp.

THE "nuclear winter" thesis advanced in 1983 by Turco, Toon, Ackerman, Pollack, and Sagan ("TTAPS")[1] seemed to portend a evolution in strategic theory and policy. One commentator declared that he nuclear winter data "change everything in the world about the prosect of thermonuclear warfare."[2] Another observer asserted that the nulear winter science "seemed to promise a fundamental change in the ay the world looked at nuclear weapons and war."[3] A reviewer of *The :old and the Dark*[4] maintained that it could be "the most important book ver published," with greater impact on the direction of humanity than the Odyssey, the Bible, the Koran, or the collected works of William hakespeare."[5]

Drawing on the perspective gained in the intervening years, this essay ims to examine the effect of the nuclear winter thesis upon the strategic ebate Our argument is that both "hawks" and "doves" responded to ne nuclear winter thesis in an ideological manner, illustrating the theo-)gical nature of the debate on nuclear strategy.

[1] Richard P. Turco, Owen B. Toon, Thomas R. Ackerman, James B. Pollack, and Carl agan (TTAPS), "Nuclear Winter: Global Consequences of Multiple Nuclear Explosions," *ience* 222 (December 23, 1983), 1283–90.

[2] Lewis Thomas, "Foreword," in Paul R. Ehrlich et al., eds., *The Cold and the Dark: The orld after Nuclear War* (New York: W. W. Norton, 1984), xxi.

[3] Gregg Herken, *Counsels of War*, expanded ed. (New York: Oxford University Press, 987), 331.

[4] Ehrlich et al. (fn. 2).

[5] Cited in Russell Seitz, "In from the Cold: 'Nuclear Winter' Melts Down," *National Interest* 5 (Fall 1986), 7.

The Nuclear Winter Science

TTAPS concluded that a 5,000-megaton exchange in which urban are
and hardened military targets were destroyed could result in a month
long period of darkness and subfreezing temperatures.[6] Only a sma
percentage of the solar radiation would penetrate the pall of dust an
smoke, and average temperatures in large portions of the Norther
Hemisphere might drop by as much as 15 to 25 degrees centigrade du
ing the summer. (A nuclear exchange in the late fall would have less
climatic effects, because sunlight and temperatures are already low, an
the growing season is over.) Effects of similar magnitude would occur
100 megatons (in the form of one thousand 100-kiloton bombs) were ai
burst in urban areas. Attacks against cities are more likely to cause
nuclear winter than are attacks against isolated military targets, becaus
of the great volume and opacity of the soot generated by burning citie
Accordingly, low-yield surface bursts directed against military target
would have less severe consequences.

Several studies prior to TTAPS attempted to estimate the conse
quences of nuclear war. In 1979 the Office of Technology Assessmen
concluded that 88 percent of the U.S. population could die as a result o
a nuclear war in which cities were targeted. Even a strike against U.S
missile silos would result in three to twenty-two million casualties.[7] A
1982 World Health Organization report concluded that survivors of a
nuclear war would face epidemics, radiation sickness, contamination o
food and water supplies, inadequate medical assistance, psychologica
disorders, food shortages, and the breakdown of social and economi
systems.[8] In 1982 Paul Crutzen and John Birks argued that the fires re-
sulting from nuclear explosions would produce tar and soot, each of
which would absorb sunlight and ash, causing "twilight at noon."[9]

The TTAPS study stimulated a great deal of additional research and
commentary on the consequences of nuclear war.[10] Some studies exam-
ined the biological and environmental consequences of subfreezing tem-
peratures, low light levels, and radiation. Paul Ehrlich and associates

[6] TTAPS (fn. 1).

[7] U.S. Congress, Office of Technology Assessment, *The Effects of Nuclear War*, OTA-NW-89 (Washington, D.C.: OTA, May 1979).

[8] World Health Organization, *Effects of Nuclear War on Health and Health Services*, Publication no. A26-12 (Rome: World Health Organization, 1982).

[9] Crutzen and Birks, "The Atmosphere after a Nuclear War: Twilight at Noon," *Ambio* 2, no. 2–3 (1982), 115–25.

[10] Reviews of the nuclear winter science are provided in Mark A. Harwell and Christine C. Harwell, "Updating the Nuclear Winter Debate," *Bulletin of the Atomic Scientists* 43 (October 1987), 42–44; and Alan Robock, "New Models Confirm Nuclear Winter," *Bulletin of the Atomic Scientists* 45 (September 1989), 32–35.

ɔncluded that "the population size of Homo sapiens conceivably could e reduced to prehistoric levels or below" in the aftermath of a 10,000-ıegaton nuclear war.[11] The Scientific Committee on the Problems of the ;nvironment of the International Scientific Union determined that a nu-lear war could result in "a potential loss of about one to a few billion umans from long-term consequences."[12]

The TTAPS study also attracted a great deal of criticism. Some sci-ntists argued that the TTAPS model overstated the consequences of uclear winter or that there remained a great deal of uncertainty as to ıe likely length of a nuclear winter. For example, George Carrier, co-hair of a 1984 National Academy of Science study on the subject, noted ıat "the uncertainties that pervade the quantitative assessment of the tmospheric effects of a major nuclear exchange are so numerous and so large that no definitive description of those effects is possible at this time."[13] Similarly, the General Accounting Office concluded in 1986 that the effect of a nuclear exchange upon the production of dust and smoke, its injection into the atmosphere, and the atmospheric response to these contaminants could not be predicted with certainty.[14]

Several analysts have criticized the nuclear war scenarios used as the baseline case by the TTAPS group. Albert Wohlstetter observed that these scenarios "hardly describe a sequence of events at all. Rather, they list a collection of simultaneous explosions by the two antagonists well-designed to bring on a global disaster."[15] Russell Seitz asserted that the TTAPS model ignored physical reality, as it featured a planet without continents or oceans, and days of twenty-four hours of sunlight at one-third strength. Instead of hypothesizing realistic smoke emissions, the TTAPS model "simply dumped a ten-mile soot cloud into the atmosphere instantly." The model also "dealt with such complications as east-west winds, sunrise, sunset, and patch clouds in a stunningly elegant manner—they were ignored."[16]

Much recent research has predicted less severe temperature reductions than those predicted by TTAPS. Starley Thompson and Stephen

[11] Ehrlich et al., "Long-Term Biological Consequences of Nuclear War," *Science* 222 (December 23, 1983), 1293.

[12] Mark A. Harwell and Thomas C. Hutchinson, *Environmental Consequences of Nuclear War*, vol. 2, *Ecological and Agricultural Effects* (New York: John Wiley, 1985), 484.

[13] Carrier, "Nuclear Winter: The State of the Science," *Issues in Science and Technology* 1 (Winter 1985), 114–17, at 117.

[14] U.S. General Accounting Office, *Nuclear Winter's Uncertainties Surrounding the Long-Term Effects of Nuclear War* (Washington, D.C.: GAO, March 1986).

[15] Wohlstetter, "Between an Unfree World and None: Increasing Our Choices," *Foreign Affairs* 63 (Summer 1985), 962–94, at 977.

[16] Seitz (fn. 5), 4.

Schneider, for example, concluded that a "nuclear fall" was a scenario than a nuclear winter, since temperature reduction mitigated by factors such as the capacity of oceans to hold h "greenhouse effect" of the smoke.[17] Other analysts have a: TTAPS overestimated the amount of combustible material in States, Europe, and the Soviet Union.[18]

In a 1990 report the TTAPS group revised its original pre 15–25 degree centigrade temperature reduction from a nucle: to one of 10–20 degrees centigrade. They point out, howeve: this relatively small decrease in temperatures could be enoug widespread crop failure, famine, and "more human casualti direct effects of nuclear war.[19]

Sagan and Turco, in *A Path Where No Man Thought*, c(original TTAPS conclusions on the effects of nuclear wint(latest findings of numerous researchers. They maintain that modest modifications, the original TTAPS calculations are sti 445–53).

THE HAWKS' RESPONSE

Many hawkish and dovish strategic analysts strove (sometime fashion) to utilize nuclear winter as an argument for their held policy preferences. For hawks, nuclear winter was an argument for counterforce and limited nuclear options; for d(an additional argument for disarmament and arms control.

One response of American hawks to the TTAPS report ' Thomas Powers reported that Pentagon officials with who versed seemed to have "a wistful hope that 'more study' [w(the nuclear winter problem go away."[20] Steven Kull condu views with numerous high-ranking American defense offici ported the feeling of some that "the apocalyptic character of n had been exaggerated." Some argued that "damage to citie: limited because the Soviets have a counterforce targeting dc

[17] Thompson and Schneider, "Nuclear Winter Reappraised," *Foreign Affair* 1985), 163–78.

[18] Malcolm W. Browne, "Nuclear Winter Theorists Pull Back," *New York T* 23, 1990, pp. C1, C8.

[19] R. P. Turco et al., "Climate and Smoke: An Appraisal of Nuclear Winte (January 12, 1990), 166–76.

[20] Powers, "Nuclear Winter and Nuclear Strategy," *Atlantic Monthly*, Nove 64.

addition, "a fair number dismissed the nuclear winter theory, though their basis for doing so was not well articulated."[21]

Another response of American hawks was to argue that nuclear winter was an additional impetus for counterforce, limited nuclear options, strategic modernization, and active and passive defense. In March 1985, at the direction of Congress, the Pentagon produced a report on the implications of the nuclear winter thesis for U.S. strategy. The report argued that if the environmental effects of nuclear war were worse than previously imagined, the case for a U.S. escalation dominance capability is even stronger: "It is our position that, however an adversary chooses to initiate nuclear conflict, we must have forces and a targeting capability so that our response would deny either motive or advantage to the aggressor in further escalating the conflict."[22]

The Pentagon report also asserted that an improved system of command, control, communications, and intelligence (C^3I) could help to control possible escalation to nuclear winter effects: "Survivable C^3I contributes to escalation control and thus . . . to mitigation of damage levels (of whatever kind, including possible climatic effects) by reducing pressures for immediate or expanded use of nuclear weapons out of fear that capability for future release will be lost."[23]

American hawks also claimed that the nuclear winter thesis should compel strategic planners to acquire weapons that would not trigger a nuclear winter. The Pentagon report made the case that MARVs, earth-penetrating warheads, and non-nuclear strategic weapons "could reduce yields and in other ways limit the starting of fires."[24] Michael Altfeld and Stephen Cimbala suggested that "targets within cities could be struck without creating global catastrophe" by using "weapons of very low yield delivered with precision accuracy."[25] Wohlstetter also endorsed the concept of environmentally sensitive weapons and targeting policies: "We should make it clear that we will not respond to an attack in any way that would have a substantial chance of causing a nuclear winter."[26] RAND analyst J. J. Gertler argued that the development of enhanced radiation weapons (ERWs) would be desirable, because "they would have much less nuclear winter effects than conventional nuclear devices of

[21] Kull, *Minds at War: Nuclear Reality and the Inner Conflicts of Defense Policymakers* (New York: Basic Books, 1988), 75.

[22] Secretary of Defense Caspar W. Weinberger, "The Potential Effects of Nuclear War on the Climate: A Report to the United States Congress," Washington, D.C., March 1985, p. 11.

[23] Ibid., 12.

[24] Ibid., 15.

[25] Altfeld and Cimbala, "Targeting for Nuclear Winter: A Speculative Essay," *Parameters* 15, no. 3 (1985), 8–15, at 11.

[26] Wohlstetter (fn. 15), 990.

similar yield." Neutron bombs would be particularly useful on Europe battlefields, "whose densely populated cities and thick forests would pı duce a heavy combustible fuel load." Furthermore, "ERW and (debatab chemical weapons are the only super-conventional systems whose effe can be restricted to theater or battlefield," and they "thus become mo important as deterrents to conventional attack."[27] Peter deLeon not that

> accelerated development and deployment of earth-penetrating and wide-area munitions [could] utilize either conventional or low-yield nucleaı warheads as a means to debilitate hardened Soviet targets (such as missile silos) without incurring the nuclear winter costs of the magnitude currently debited to larger warheads.[28]

In addition, Robert Kennedy of the Army War College stated that c fense planning should move toward a reduction in warhead yield a the deployment of less vulnerable mobile missiles.[29]

American hawks also argued that the nuclear winter thesis shou lead policymakers to emphasize limited nuclear options and selective ta geting. While some supporters of assured destruction argued that t flammability of hostage cities greatly enhances deterrence, hawks mai tained that intentionally targeting cities is both immoral and strategica unwise. The Pentagon's nuclear winter report stated that counterval strikes "necessarily have a high probability of starting major fires aı consequently, of creating large amounts of smoke."[30] For Colin Gra the nuclear winter thesis should assist those "still in need of enlighte ment on nuclear strategy," by serving as "the final nail in the coffin the concept of mutual assured destruction (MAD) and its variants,' which, according to Gray, becomes even less credible given the inhe ently catastrophic nature of its implementation. While some doves a serted that the nuclear winter thesis invalidated scenarios of a limit nuclear war, hawks argued instead that the thesis reinforced incentiv to control escalation and war termination.[32]

[27] Gertler, *Some Policy Implications of Nuclear Winter*, P-7045 (Washington, D.C., RAN Corporation, January 1985), 10.

[28] DeLeon, "Rethinking Nuclear Conflict: Strategic Implications of a Nuclear Winte *Defense Analysis* 3 (December 1987), 319–36, at 331.

[29] Robert Kennedy, "Nuclear Winter, War Prevention, and the Nuclear Deterrent," Sederberg, 160–61. It should be noted that these policy prescriptions have not been adopt by the Pentagon, which still seeks large warheads such as the MX, and whose targeting pla still call for attacking cities. See Lawrence W. Woodruff, "Nuclear Force Modernizatio Efforts," *Defense Issues*, March 11, 1988, pp. 1–8.

[30] Weinberger (fn. 22), 11.

[31] Gray, "The Nuclear Winter Thesis and the U.S. Strategic Policy," *Washington Quarte* 8 (Summer 1985), 85–96, at 94.

[32] Fred J. Reule, *Nuclear Winter: Asymmetrical Problems and Unilateral Solutions*, Airpow

One reason that the nuclear winter thesis adds to the case for limited ıuclear options, according to American hawks, is that the Soviets are ikely to emphasize their own capability for selective strikes. Wohlstetter tated that a Soviet military planner might seek to maximize the nuclear vinter effect by targeting cities and other targets with high densities of uel, as well as by exploding "high-yield weapons at altitudes that would naximize the thermal pulse over combustible areas and so send smoke n huge quantities into the atmosphere." However, Wohlstetter finds it 'far more likely" that a Soviet military planner "might use nuclear veapons to accomplish some military purpose in the course of war, and lo it in a way that takes account of the fact that destruction extraneous o that purpose would cause a nuclear winter and make that military ›urpose idle."[33] Altfeld and Cimbala maintained that the effect of nu:lear winter "may well be to make nuclear war appear much more wageıble and controllable than it now appears to be, and deceptively similar o past wars," due to incentives for city avoidance and the acquisition of ow-yield, high-accuracy weapons.[34]

RAND analyst Philip Romero argued that nuclear winter makes nuclear war more likely. He reasoned as follows: (1) both superpowers will attempt to minimize the nuclear winter effect by deploying low-yield, high-accuracy weapons; (2) this will make concealment and mobility the only means of protecting one's arsenal; (3) the pace of a war would then depend on the ability of each side to acquire information on the location of its opponent's weapons; and (4) both sides will seek to destroy each other's reconnaissance satellites and intelligence centers. Therefore, Romero concluded, "attacks will probably be very small (one to perhaps ten weapons per volley). A true 'slow-motion' war could last months or years."[35]

American hawks also used the nuclear winter thesis as an argument for strategic defense, since a successful defense of cities would reduce the amount of smoke injected into the atmosphere. President Reagan stated in 1985 that a nuclear war "could end up in a victory for no one because we would wipe out the earth as we know it." Therefore, in order to avoid "the nuclear winter that scientists have been talking about," a ballistic missile defense system should be constructed "in such a way that

Research Institute, Maxwell Air Force Base, Alabama, Report no. US-ARI-CP-86-1, September 1986, p. 27.

[33] Wohlstetter (fn. 15), 972.

[34] Altfeld and Cimbala (fn. 25), 13.

[35] Romero, *Nuclear Winter: Implications for U.S. and Soviet Strategy*, P-7009 (Washington, D.C.: RAND Corporation, August 1984), 18.

those missiles aren't going to get to their target."[36] The Pentagon's report on nuclear winter stated that a defensive system "may provide a greater mitigating effect on atmospheric consequences" than likely arms reduction agreements.[37]

The policy response of American hawks to the nuclear winter thesis has numerous defects. The Pentagon's proposed environmentally sensitive policies would prove ineffective unless the Soviets and third parties follow suit—a most unlikely result. The Soviet nuclear arsenal is full of "heavy" missiles, French strategic policy is based on the idea of countervalue targeting, and decision makers in N + 1 countries are not likely to be interested in restricting the megatonnage of their arsenals. Moreover, a nuclear war that started with "technical fix" weapons might not end with them unless all sides retired their heavy missiles and/or exercised incredible self-restraint.

The hawks' analysis of how nuclear war should be conducted in light of the nuclear winter thesis is also misguided. Romero's scenario for a protracted, intermittent nuclear war is fanciful. What he failed to explain is why the United States and the Soviet Union would continue a war in such "slow motion." Clearly, these strikes would be of limited military utility, since a salvo would probably expend more warheads than it would destroy. Furthermore, national leaders would have ample time to negotiate a settlement between the rounds of a war of attrition. Alternatively (also in opposition to Romero's view), an initial strike could escalate into all-out destruction, due to the nature of decision making in times of crisis and the colocation of civilian and military targets.[38]

The hawks' analysis also appears to have ignored the fact that the effects of nuclear winter could seriously degrade the performance of command and control facilities. The director of the Navy's Strategic and Theater Nuclear Warfare Division speculated that the dust raised by nuclear detonation could also seriously damage bombers, cruise missiles, and airborne command posts.[39] In short, the postattack environment does not appear to be conducive to the execution of a carefully orchestrated multiple-exchange nuclear war.[40]

[36] "Transcript of Interview with President on a Range of Issues," *New York Times*, February 12, 1985, p. A10.

[37] Weinberger (fn. 22), 13.

[38] Desmond Ball, *Can Nuclear War Be Controlled?* Adelphi Paper no. 169 (London: International Institute for Strategic Studies, 1981).

[39] R. F. Bacon, U.S. Congress, House Committee on Science and Technology, Hearing, *The Climatic, Biological, and Strategic Effects of Nuclear War*, September 22, 1984, p. 231.

[40] Palomar Corporation (Washington, D.C.), "Implications of the 'Nuclear Winter' Thesis" (Prepared for the Defense Nuclear Agency, Contract #001-84-C-0257, June 25, 1985), 22–23.

Finally, all of the measures advocated by the hawks in response to the nuclear winter thesis—escalation dominance, command and control improvements, low-yield weapons, selective nuclear targeting, and strategic defenses—all fall prey to the traditional damage limitation/war prevention trade-off. Strategic analysts have long faced the "usability paradox"—that making nuclear weapons too destructive lessens the credibility of their use, while making them less destructive may lessen constraints against their use.

The Doves' Response

American "doves," of course, had a vastly different policy response to the nuclear winter thesis. They suggested that nuclear winter reinforced the need for one of three policies: mutual assured destruction, arms control, or global disarmament.

Some doves embraced the nuclear winter concept, arguing that it buttressed the fact that mutual assured destruction is an inescapable physical reality rather than a mutable strategic doctrine. Former Admiral Noel Gayler posited that nuclear winter makes deterrence much easier, because "there will be no bystanders" in a nuclear war, because the nuclear balance becomes meaningless if only a relatively small number of warheads is required to trigger a nuclear winter, and because cheating on agreements would have to be immense to make a difference. For Gayler, "the policy implications of this analysis are straightforward. The prospect of nuclear winter reinforces the already totally unacceptable character of nuclear war."[41]

Similarly, Herbert Simon declared that "the nuclear winter scenario replaces the prospect of mutual destruction through the failure of mutual deterrence by the prospect of assured self-destruction through nuclear attack." Given nuclear winter, "a second strike is no more appetizing than a first, and a preemptive strike loses any gambling appeal that it might have had."[42]

George Quester made the argument that nuclear winter reinforces the logic of mutual assured destruction, because it increases the costs of initiating a nuclear war. The nuclear winter theory answers "with a resounding reassurance" that retaliation against a nuclear first strike "will be adequate—as long as the two nuclear stockpiles remain so large." Deterrence is assured "because the Soviets will starve if they apply a

[41] Admiral Noel Gayler, U.S. Congress, Joint Economic Committee, Hearing, *The Consequences of Nuclear War*, July 11 and 12, 1984, p. 6.

[42] Simon, "Mutual Deterrence or Nuclear Suicide," *Science* 223 (February 24, 1984), 775.

thermonuclear torch to U.S. cities." Nuclear winter provides "a reassurance of deterrence" which is "altogether satisfying and convincing." Quester concluded that the "ultimate moral response" to the nuclear question remained the doctrine of mutual assured destruction, as understanding "how bad a nuclear war would be keeps national leaders from crossing the line into such a war."[43]

For Dan Horowitz and Robert Lieber, "the nuclear winter phenomenon could require significant revisions in prevailing assumptions about nuclear deterrence." They assert that since nuclear war "could bring about the devastation or even extinction of human civilization," both disarming first strikes and nuclear war-fighting doctrines become less credible.[44]

Similarly, Carl Sagan argued in his essay in *The Long Darkness* that the nuclear winter thesis invalidated first strike scenarios and theories of limited nuclear war. With nuclear winter, the "retaliation" of the victimized nation would be automatic, and a first strike would be suicidal, as it would create a Pyrrhic victory of only ten days' duration "before the prevailing winds carry the nuclear winter to the aggressor nation." The "attractiveness" of the first strike "would seem to be diminished significantly" (Grinspoon, 38).

Sagan recognized that many hawks advocated the development of environmentally sensitive weapons to attempt to circumvent the effects of nuclear winter. Sagan opposed such a policy, since development of "post-TTAPS first strike weapons" could denote "a serious interest in making a climatically safe but disabling first strike." He also asserted that such an effort would be futile, as limited nuclear war would inevitably escalate to all-out war (Grinspoon, 42).

The nuclear winter thesis solidified the doves' opposition to ballistic missile defense. If the effects of a nuclear exchange are larger than previously anticipated, defense against incoming missiles becomes even more difficult. Even a ballistic missile defense system with a very high degree of success (even 90 percent) would be unsatisfactory, as it might lead to a false sense of confidence. A ballistic missile defense system could never protect American cities, because any attempt to do so would "likely" cause the Soviet Union "to target its missiles so as to maximize damage to the U.S. population," in spite of the fact that this would "pose

[43] Quester, "Nuclear Winter: Bad News, No News, or Good News?" in Catherine McArdle Kelleher, Frank J. Kerr, and George H. Quester, eds., *Nuclear Deterrence: New Risks, New Opportunities* (Washington, D.C.: Pergamon-Brassey's, 1986), 78–79, 80.

[44] Horowitz and Lieber, "Nuclear Winter and the Future of Deterrence," *Washington Quarterly* 8 (Summer 1985), 59–70, at 68.

a serious danger of triggering a climatic catastrophe."[45] The Soviet Union could also attempt to overcome the system with a proliferation of nuclear warheads (Sagan, in Grinspoon, 45).

Doves also made the point that a terminal-phased interception system would contribute to a nuclear winter, since nuclear-tipped missiles that detonated at very low altitudes would start fires. This problem could be avoided with a boost-phase or non-nuclear defensive system. However, defensive systems would have a difficult time functioning in a dust- and smoke-laden environment. Sensors would have to be able to discriminate between warheads and decoys. One possible system would use long-wave infrared radiation (LWIR) to perform this task, "by trying to observe small differences in intensity, wavelength, and object motion." Unfortunately, the LWIR signal of an object "will also vary with the altitude of injected smoke, the size of the smoke particles, and their optical properties, as well as with the temperature of the upper atmosphere." Therefore, a defensive system "must have a very detailed understanding of the physical environment in which its sensors operate."[46] In short, nuclear winter would add to the "fog of war."[47]

While some doves stressed the contribution of the nuclear winter thesis to deterrence, others argued that nuclear winter made arms control an even greater imperative. Doves proposed U.S.-Soviet confidence-building measures, actions to reduce the risk of accident or miscalculation, the de-MIRVing of missiles, the abandonment of plans for limited nuclear war, retiring old weapons systems with high yields, improving communications at all levels, ratifying the SALT II Treaty, banning all nuclear warheads in excess of 300 kilotons, and reaching agreement with the Soviets to not target cities.

The nuclear winter thesis was also used by doves to argue for a doctrine of "no first use." If "in the age of massive thermonuclear overkill it no longer makes any sense—if it ever did—to hold these weapons for any other purpose than the prevention of their use," then the execution of a first use threat would be suicidal.[48] An American nuclear response to a Soviet invasion of Western Europe would be especially illogical if it were to result in the destruction of the American homeland even without a Soviet response. Bernard Weissbroud asserted that the nuclear winter

[45] Carl Sagan, U.S. Senate, Committee on Armed Services, Hearing, *Nuclear Winter and Its Implications*, October 2 and 3, 1985, p. 138.

[46] Theodore Postol (fn. 39), 104, 101.

[47] Herman Kahn, *On Escalation: Metaphors and Scenarios* (reprint, Westport, Conn.: Greenwood Press, 1986).

[48] McGeorge Bundy et al., "Nuclear Weapons and the Atlantic Alliance," *Foreign Affairs* 60 (Spring 1982), 753–68, at 768.

thesis makes "providing an adequate conventional defense of Europe a matter of the utmost urgency.[49]

Many doves argued that the nuclear winter science created an additional imperative for nuclear disarmament. Helen Caldicott claimed that since "we hold in our hands the ability to destroy most of creation," mankind must rid the earth of this threat.[50] Herbert Simon maintained that because of nuclear winter, "the way is now open for nuclear disarmament, even partial unilateral nuclear disarmament."[51] The authors of the Soviet publication *Global Climatic Catastrophes* concluded that the best way to avoid "the unpremeditated self-destruction of mankind" would be "a gradual reduction and final elimination of all types of weapons that threaten the fate of mankind."[52] Three British authors contended that the only solution to the prospect of nuclear winter is "to halt and reverse the nuclear arms race."[53] David E. Fisher deduces that "since the idea of our launching fifteen thousand nuclear warheads is no longer feasible, there's not much point in keeping those fifteen thousand warheads ready to fly at the touch of a button" (pp. 203–4). Carl Sagan called for the superpowers to reduce their nuclear arsenals below the nuclear winter threshold, as a matter of "elementary planetary hygiene," because "it seems clear that the species is in grave danger at least until the world arsenals are reduced below the threshold for climatic catastrophe" (Grinspoon, 44).

Another argument that is consistent with the dovish position is that nuclear winter could be used to strengthen deterrence by creating a "doomsday machine" (Sederberg, 10).[54] The objective of a doomsday machine would be to convince potential aggressors that any nuclear conflict would bring about the destruction of the aggressor state (as well as the rest of humanity). A doomsday machine could be created by intentionally locating military assets in or near cities, by deploying mobile missiles (which could be moved near cities in a crisis), by deploying missiles in multiple protective shelters (forcing aggressors to launch a barrage attack that would induce nuclear winter), or by using "dirty bombs" (which could render cities uninhabitable with radioactive fallout).[55]

[49] Weissbroud, "Are Nuclear Weapons Obsolete?" *Bulletin of the Atomic Scientists* 40 (August–September 1984), 8.

[50] Caldicott, *Missile Envy* (New York: Bantam Books, 1984), 307.

[51] Simon (fn. 42), 775.

[52] Michael I. Budyko, Georgia S. Golitsyn, Yuri A. Izrael, *Global Climatic Catastrophes*, trans. V. G. Yanuta (Berlin: Springer-Verlag, 1988), 85.

[53] Owen Greene, Ian Percival, and Irene Ridge, *Nuclear Winter: The Evidence and the Risks* (Cambridge: Polity Press, 1985), 19.

[54] The concept of a doomsday machine originated with Herman Kahn, *On Thermonuclear War* (Princeton: Princeton University Press, 1960).

[55] Romero (fn. 35), 13, 16. Romero points out that Soviet war planners would have a more difficult time employing "dirty bombs" because of American urban sprawl—a "suburb gap."

Sagan and Turco's recent book on nuclear winter, *A Path Where No Man Thought*, provides a comprehensive overview of the dovish view on nuclear winter and is thus deserving of special attention. While Sagan and Turco call for such measures as missile de-MIRVing, phasing out "short-flight-time intermediate range delivery systems in Europe and Asia," "gradual elimination of all forward-based tactical nuclear forces," conventional force reductions, and abandonment of the Strategic Defense Initiative (pp. 228–29), their clear policy preference is "minimum sufficient deterrence," to be followed by nuclear abolition.

Sagan and Turco maintain that minimum sufficient deterrence, defined as the "few hundred strategic weapons" that are "clearly enough to obliterate the United States or the Soviet Union as functioning economic and political entities," would provide "*greater* strategic stability than the present arsenals, and constitute a robust deterrent to conventional attack (p. 135; emphasis in original). For Sagan and Turco, "the specific nuclear winter objective in the transition to minimum sufficiency is to *generate a stable deterrence in which the probability of nuclear war is extremely small, but where, if the worst happens, there would be no significant global climatic consequences* (p. 233; emphasis in original). They maintain that there should not be "a hint of a Doomsday Machine" in global nuclear arsenals (p. 219). However, minimum sufficient deterrence is but "a way station to abolition—from which we can safely survey the landscape and the remaining obstacles" (p. 299).[56]

Sagan and Turco also address the problem of nuclear proliferation in a "minimum sufficient deterrence" regime. Their message is that if the superpowers set an example of disarmament, other states will fall in line, particularly because "the resources available to the superpowers" are "so great" that "adequate incentives could doubtless be found even for the most recalcitrant state," including "Israel, Iraq, South Africa, India, Pakistan, North Korea, Brazil, and Argentina (p. 286). While export controls would help to prevent proliferation, what is fundamentally required is "arms reductions from the top—those nations that first established the arms race gear-train, that *drive* the gear-train, have the responsibility, and increasingly the incentive, to stop the engine and demolish the links (p. 290; emphasis in original). Sagan and Turco claim that Britain, France, and China are willing to agree to minimum deterrence (p. 134), that Israel may have acquired nuclear weapons only because of the presence of American and Soviet nuclear weapons in their Mediterranean and Persian Gulf fleets, and that "withdrawal of Ameri-

[56] Sagan and Turco also state that if nuclear winter research indicates the "threshold" at about one hundred weapons each, "nuclear winter would then suggest that we move quickly from minimum sufficiency to abolition" (p. 416).

can and Soviet nuclear weapons from the Middle East—which would be a natural consequence of an MSD regime—might assist Israel in reducing its arsenals" (p. 291).

One reason that Sagan and Turco are so optimistic about the possibility of minimum deterrence and nuclear abolition is that "an era of improved relations between the United States and the Soviet Union is the optimum time to reassess military doctrine and policy, to reconsider weapons systems on order, to reverse the arms race (pp. xviii–xix). "Astonishingly," they write, the world is witnessing "conventional force reductions and major advances in international amity" (p. 286). Prospects for future international amity are even brighter, they state, because of "global economic interdependence, the unexpected opening up of the Soviet Union and Eastern Europe to the rest of the world, the merging European Union, the improving effectiveness and acceptability of the United Nations and the World Court, and the growing political success of the environmental agendas," as well as improved global communications through television, telephones, facsimile machines, and computer networks (p. 300).

Nevertheless, Sagan and Turco warn against excessive optimism, arguing that it is a "dangerous illusion [to] think of the problem of nuclear war as solved" (p. xviii). Despite "all the genuine goodwill in the present attitudes of the superpowers and the profound changes in their relationship," the fact remains that each superpower has over 10,000 nuclear warheads, aimed "with fine premeditation" at specific targets on the other side (p. xviii). They also note that "there are over 10,000 tactical warheads and nuclear 'munitions' in Western and Eastern Europe, apart from the USSR" (p. 266).

The doves' analysis of the policy implications of the nuclear winter thesis is in numerous ways unpersuasive. Initially, the doves' argument that even a very successful ballistic missile defense system would have little value in preventing a nuclear winter is faulty. Even a "leaky" defensive system might prevent nuclear winter effects if arsenals are reduced significantly, and the dovish analysis makes the unprovable assumption that the superpowers would use their entire arsenal in a nuclear war and that cities would be the prime targets. Also, given the apparent scientific consensus that no single nuclear winter threshold exists, and that the nuclear winter is better described as a "progressive set of increasingly nasty ecological effects,"[57] it is possible that a defensive

[57] Donald M. Snow, "Strategic Uncertainty and the Nuclear Winter: Implications for Policy," in Paul Viotti, ed., *Conflict and Arms Control: An Uncertain Agenda* (Boulder, Colo.: Westview Press, 1986), 63.

system could make a meaningful difference in the length and severity of a nuclear winter.

Further, Sagan and Turco's recommendation of a move toward minimum deterrence and nuclear abolition is both undesirable and unrealistic. Minimum deterrence would leave the nuclear balance far less stable, by making asymmetries in nuclear capabilities much more significant than under levels of high force. Minimum deterrence would also make extended deterrence on the part of the superpowers virtually impossible—thereby eliminating one of the strongest proliferation antidotes available. Very low levels of nuclear force among the major powers would also increase the attractiveness to LDCs of membership in the nuclear club and could make the world safe for conventional war, or even limited nuclear war (an outcome that Sagan and Turco even appear to *endorse*; p. 233). In addition, Sagan and Turco offer no insight as to how an international agreement requiring minimum deterrence or nuclear abolition is to be achieved or enforced.

Sagan and Turco's position on force levels is somewhat mystifying. They apparently believe that the fear of nuclear winter is an important contributor to the "end of the arms race," yet the policy they advocate would lessen the deterrence added by nuclear winter effects; indeed, it would create incentives for arms racing.[58] Given the fact that nuclear weapons may have created—or at least created the conditions for—the "long peace" of the postwar period,[59] it is up to advocates of radical policy change to demonstrate that their proposals would not disturb international stability.

There is also very little credibility to Sagan and Turco's analysis of the willingness of the international community to disarm. Nuclear weapons are central elements of the defense policies of Great Britain, France, and China. Statements quoted by Sagan and Turco about the willingness of those countries to move toward minimum deterrence and nuclear abolition should be read as merely rhetorical, belied by decades of actual behavior. Sagan and Turco's justification of past Israeli behavior, as well as their predictions of future Israeli willingness to disarm, are hardly credible, as the Israeli government has made it quite clear that it is not interested in disarming and that it is quite prepared to use military force. Furthermore, Sagan and Turco's optimism about the ability of the superpowers to influence the nuclear decisions of such states as Iraq and

[58] McGeorge Bundy, "The Bishops and the Bomb," *New York Review of Books*, June 16, 1983.

[59] John Lewis Gaddis, "The Long Peace: Elements of Stability in the Postwar International System," *International Security* 10 (Spring 1986), 99–142.

North Korea is remarkable, as is their failure to specify just how this to be accomplished.

Sagan and Turco are also guilty of taking contradictory positions c the impact of the nuclear winter thesis on deterrence. At one point th state that "nuclear winter strengthens deterrence" because it makes possible that a massive first strike, "intended to destroy the retaliato forces of the adversary, could boomerang and destroy the attackers wit out any effort on the part of the victim" (p. 145). However, at anoth point they state that "nuclear winter does, though, tend to undermi the current doctrine of deterrence," because with nearly sixty thousan nuclear weapons in the world, "how could the threat of a massive fir strike or of massive retaliation be believable, if it inexorably puts th attacker in mortal peril?" (p. 224).

Sagan and Turco, noting that the world has recently undergon "astonishing" changes, conclude that international comity, aided by ec nomic interdependence and international communication, have ma nuclear abolition possible. Sagan and Turco appear to belong at the "li eral 'end of history' optimism" camp,[60] yet they show no awareness this recent debate.[61] The outbreak of war in the Persian Gulf adds litt to the credibility of Sagan and Turco's optimism.

While on one level Sagan and Turco believe that everything in t world has changed, on another level they believe that absolutely nothi has changed. Just like Richard Cheney[62] and Colin Gray,[63] they mainta

[60] Jack Snyder divides the debate over the future of Europe into three views: "liberal 'e of history' optimism," "Hobbesian pessimism," and "neo-liberal institutionalism"; Snyd "Averting Anarchy in the New Europe," *International Security* 14 (Spring 1990), 5–41.

[61] Prominent proponents of liberal optimism include Francis Fukuyama, "The End History?" *National Interest* 16 (Summer 1989), 3–18; Michael Doyle, "Liberalism and Wor Politics," *American Political Science Review* 80 (December 1986), 1151–69; idem, "Kant, Li eral Fallacies, and Foreign Affairs, Part I," *Philosophy and Public Affairs* 12 (Summer 198 2205–35; idem, "Kant, Liberal Fallacies, and Foreign Affairs, Part II," *Philosophy and Publ Affairs* 12 (Fall 1983), 323–53; John Mueller, *Retreat from Doomsday: The Obsolescence of Maj War* (New York: Basic Books, 1989); James Lee Ray, "The Abolition of Slavery and the En of International War," *International Organization* 43 (Summer 1989), 405–39; and Sissela Bo *A Strategy for Peace: Human Values and the Threat of War* (New York: Pantheon Books, 1989 Prominent critics of this view include John J. Mearsheimer, "Back to the Future: Instabilit in Europe after the Cold War," *International Security* 15 (Summer 1990), 5–56; idem, "Wh We Will Soon Miss the Cold War," *Atlantic Monthly*, August 1990, pp. 35–37, 40–42, 44–4 50; Samuel P. Huntington, "No Exit: The Errors of Endism," *National Interest* 17 (Fall 1989 3–11; Leon Wieseltier, "Spoilers at the Party," *National Interest* 17 (Fall 1989), 12–16; an Richard N. Gardner, "The Comeback of Liberal Internationalism," *Washington Quarterly* 1 (Summer 1990), 23–39.

[62] Cheney stated that despite "Gorbachev's stated commitment to a defensive military doc trine, the Soviet Union continues upgrading its forces and improving their capabilities" U.S. Department of Defense, *Soviet Military Power: An Assessment of the Threat, 1988* (Wash ington, D.C.: Government Printing Office, 1988), preface, no page.

[63] Gray, *War, Peace, and Victory: Strategy and Statecraft for the Next Century* (New York

ıat dramatically improved U.S.-Soviet relations do not alter the fact that oth superpowers still aim thousands of nuclear warheads at each other. Vhat Sagan and Turco fail to consider is that the nuclear arsenals of ther countries must be viewed in their political contexts. A nuclear-rmed, democratic, market-oriented Soviet Union with a foreign policy ommitted to international peace and international law and a defense olicy strictly limited to defense is different from a nuclear-armed, non-emocratic, command economy Soviet Union with a foreign policy com-nitted to hegemonic domination and a defense policy oriented toward uclear preemption. The United States worries little about a French or British nuclear arsenal of several hundred warheads; an Iraqi or Libyan uclear arsenal of several hundred warheads is another matter entirely.

Finally, Sagan and Turco's apolitical, atheoretical approach, and their ixation on the physical effects of nuclear weapons, causes them to ignore he most significant issue facing nuclear strategists and policymakers: ow to improve international security in the post–cold war environment. Sagan and Turco appear to believe that quantifying the precise effects of uclear explosions is more important than considering Europe's security rchitecture, German unification, nuclear proliferation (especially in the Middle East), chemical weapons, the security of command and control, he possibility of nuclear accidents, changing Soviet military doctrine, ow-intensity conflict, the emergence of new geopolitical powers, the fu-ure of the Strategic Defense Initiative, drugs, pollution, population, the estructuring of NATO, the collapse of the Warsaw Pact, the possible Bal-anization of the Soviet Union, and the effects of the INF and START trea-ies.[64] Sagan and Turco's analysis, completed in the summer of 1990, eems almost to exist in a time warp.

The Pragmatic Center

t is clear from the debate over the nuclear winter thesis that numerous awks and doves have failed to come to terms with the nuclear revolu-

imon and Schuster, 1990), 14: "For the broad purposes of this work on strategy, it really oes not matter who is up or down in Moscow on the date of publication."

[64] Prominent literature on the emerging security environment includes Gary L. Guertner, *Deterrence and Defense in a Post-Nuclear World* (New York: St. Martin's Press, 1990); James . George, *The New Nuclear Rules: Strategy and Arms Control after INF and START* (New York: St. Martin's Press, 1990); Gary L. Guertner, "Strategic Arms after Reagan: The Un-inished Agenda," *SAIS Review* 10 (Winter–Spring 1990), 87–100; Mackubin Thomas Owens, "Force Planning in an Era of Uncertainty," *Strategic Review* 18 (Spring 1990), 9–22; Thomas H. Etzold, "The Strategic Environment of the Twenty-first Century: Alternative Futures of Strategic Planners," *Strategic Review* 18 (Spring 1990), 23–31; and Fred Charles Ikle and Albert Wohlstetter, *Discriminate Deterrence: Report of the Commission on Integrated Long-Term Strategy* (Washington, D.C.: U.S. Department of Defense, 1988).

tion. Certain hawks exhibit excessive zeal for nuclear war-fighting doctrines; certain doves seemingly fail to realize that nuclear weapons are not going to go away.

A third group of strategic analysts, however, has demonstrated a more sophisticated understanding of nuclear weapons and nuclear winter. What Kosta Tsipis calls "the pragmatic center" rejects both the anti-Sovietism of the Right and the utopianism of the Left. The pragmatic Center believes that since conflict between the United States and the Soviet Union cannot ever be fully resolved in an anarchic international environment, statesmen should seek international stability and predictability.[65] Those called "owls" by Graham Allison, Albert Carnesale, and Joseph Nye emphasize enhancing crisis stability, reducing the impact of accidents, developing procedures for war termination, preventing and managing crises, invigorating nonproliferation efforts, and limiting perceptions.[66]

Several strategists of the pragmatic Center analyzed the nuclear winter thesis. Nye argued that nuclear winter should stimulate additional concern about command and control capabilities, since the consequences of accidents would be greater than previously anticipated. He also urged an improvement in NATO's conventional defense capabilities, as well as serious consideration of a "no cities" targeting doctrine. Finally, he claimed that nuclear winter lessens the case for offensive force reductions, since "the benefits of restricting the strategic arsenals to low levels might be partly illusory if it is true that 100 megatons on each side targeted against cities might bring on nuclear winter."[67]

Several scholars at the Airpower Research Institute at Maxwell Air Force Base conducted a similarly impressive analysis of the effect of the nuclear winter thesis upon nuclear strategy. These analysts examined the implications of four scenarios: a low-level nuclear winter threshold, a mid-level threshold, a high threshold, and a threshold that is forever uncertain.[68] The Air Force scholars noted that the significance of the

[65] Tsipis, "The Uncommon Ground of Security Debate," *Bulletin of the Atomic Scientists* 43 (July–August 1987), 10–11.

[66] Nye, Allison, and Carnesale, "Analytical Conclusions: Hawks, Doves, and Owls," in Allison, Carnesale, and Nye, eds., *Hawks, Doves and Owls: An Agenda for Avoiding Nuclear War* (New York: W. W. Norton, 1985), 230–42.

[67] Nye, *Nuclear Ethics* (New York: Free Press, 1986), 108–13. See also idem, "Nuclear Winter and Policy Choices," *Survival* 28 (March–April 1986), 119–27.

[68] Dennis M. Drew, *Nuclear Winter and National Security: Implications for Future Policy*, Airpower Research Institute, Maxwell Air Force Base, Alabama, Report no. AU-ARI-CP-86-1, July 1986. Desmond Ball argued that the "uncertainty" scenario was the most likely outcome; see Ball, "Toward a Critique of Strategic Nuclear Targeting," in Desmond Ball and Jeffrey Richelson, eds., *Strategic Nuclear Targeting* (Ithaca, N.Y.: Cornell University Press, 1986), 32.

nuclear winter thesis varies in inverse relation to the threshold level and that the outcome of the uncertainty scenario depends on the strategic beliefs of decision makers.

Another interesting analysis was undertaken by Allen Lynch of the Institute for East-West Security Studies. Like the doves, Lynch argues, in *Political and Military Implications of the "Nuclear Winter" Theory*, that nuclear winter "reinforces the basic truth of the nuclear age: that nuclear weapons have no useful military function except to deter others from using theirs" (p. 10). Lynch also argues, with the doves, that nuclear winter "highlights the incredibility of an American nuclear guarantee to Western Europe" (p. 18).

Much of Lynch's analysis, though, defies categorization as "hawkish" or "dovish." In the area of targeting, he advocates a doctrine "that eschews both 'counterforce' and 'countervalue' strategies, one in which the non-nuclear armed forces of the opposing side become the only legitimate target of military forces, nuclear or otherwise" (p. 38). A "countercombant doctrine"[69] could be accomplished through "de-MIRVing" (reducing the ratio of warheads to missiles), "the measured introduction of strategic defensive systems to protect strategic bases," and a doctrine of no first use. These measures, he argues, would improve strategic stability by protecting the invulnerability of strategic nuclear forces.

Writing before the end of the cold war, Lynch also put forth an extremely interesting analysis of the impact of the nuclear winter thesis upon the geopolitics of Europe. He speculated that nuclear winter might bring about a radical denuclearization of Europe, which might restore Eastern Europe's role as a buffer zone for the Soviet Union and cause the Soviet Union to tighten its grip over the area (p. 35). Furthermore, the denuclearization of Europe would probably harm the American position in Western Europe, as NATO members could come to doubt the credibility of an American first use pledge that might inflict nuclear winter (p. 34).

The nuclear winter analysis of the pragmatic Center is more subtle, unpredictable, and reasoned than much of the ideological analysis of the Left and the Right. "Owlish" analysis avoids the traps of viewing nuclear weapons either as similar to conventional weapons or as weapons subject to banishment. The owlish analysis examined here also avoids the tendency of some hawks and doves to treat issues related to nuclear weapons in an apolitical manner.

[69] Lynch cites Earl Ravenal, *Strategic Disengagement and World Peace* (San Francisco: CATO Institute, 1979), 33–39.

Conclusion

The prediction of some that the nuclear winter thesis would revolutionize strategic theory and policy has not been realized.[70] The best explanation for this outcome is the fact that hawks and doves have vastly different views on nuclear strategy and, like all individuals, they seek to maintain cognitive and ideological consistency.

Robert Jervis explains that "the process of drawing inferences in light of logic and past experience that produces rational cognitive consistency also causes people to fit incoming information into pre-existing beliefs and to perceive what they expect to be there." The familiar process of constructing cognitive maps causes people to "ignore information that does not fit" and/or to "twist it so that it confirms" existing beliefs. Jervis notes that "confirming evidence, by contrast, is quickly and accurately noted."[71] Steven Kull notes that if emotions do not conform to reality assessment, individuals typically resolve the conflict by "(1) *suppressing* or ignoring conditions that suggest their emotionally gratifying behavior is not appropriate, or (2) developing complex *rationalizations* that explain how their emotionally gratifying behavior is objectively appropriate."[72] Jervis observes that since "deterrence posits a psychological relationship," it is "strange that most analyses of it have ignored decision makers' emotions, perceptions, and calculations."[73]

Hawks and doves are deeply divided in their emotions, perceptions, and calculations. They differ on a wide range of issues: whether a posture of assured destruction is adequate to deter the Soviet Union; whether the relative strategic balance has military or political significance; whether limited nuclear options make nuclear war more or less likely; whether a nuclear war could be controlled; whether Soviet strategic doctrine is oriented toward preemption; whether military technology is a cause of or a solution to the problems of the nuclear era; and whether strategic force modernization, low-yield weapons, and ballistic

[70] Indeed, several recent, notable books on nuclear strategy and defense policy fail to address the issue of nuclear winter at all. See Robert Jervis, *The Meaning of the Nuclear Revolution: Statecraft and the Prospect of Armageddon* (Ithaca, N.Y.: Cornell University Press, 1989); Joseph S. Kruzel, ed., *American Defense Annual, 1989–1990* (Lexington, Mass.: D. C. Heath, 1989); Janne E. Nolan, *Guardians of the Arsenal: The Politics of Nuclear Strategy* (New York: Basic Books, 1989); Scott D. Sagan, *Moving Targets: Nuclear Strategy and National Security* (Princeton: Princeton University Press, 1989); and Caspar W. Weinberger, *Fighting for Peace: Seven Critical Years in the Pentagon* (New York: Warner Books, 1990).

[71] Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976), 143.

[72] Kull (fn. 21), 304; emphasis in original.

[73] Jervis, "Introduction: Approach and Assumptions," in Robert Jervis, Richard Ned Lebow, and Janice Gross Stein, eds., *Psychology and Deterrence* (Baltimore, Md.: Johns Hopkins University Press, 1985), 1; emphasis in original.

lefense would contribute to strategic stability. The fundamental ı dividing hawks and doves is whether nuclear weapons have onized military strategy and world politics. Hawks believe that weapons have not eliminated the need to think about the unle and that following dovish policy advice would result in stralexibility and meaningful Soviet strategic superiority. Doves conat hawks are guilty of "conventionalizing" nuclear weapons, ould make nuclear war more likely.

lebate over the nuclear winter thesis has reflected these ideolog-rences. Hawks have argued for limited nuclear options, strategic zation, environmentally sensitive weapons, and strategic defense of their core belief that relying on a policy of assured destruction vent deterrence fails is inflexible and incredible. Doves have ar- policies of mutual assured destruction, arms control, and/or isarmament because of their core belief that nuclear war is unole and inherently uncontrollable.

ps the debate over the nuclear winter thesis is not surprising, e nature of nuclear strategy. Colin Gray observes that there is no ıc method" to nuclear strategy,[74] that there is a strategic consennly the most general of assumptions,[75] and that no one "really the role that nuclear weapons would play in war.[76]

that no one "really knows" and that nuclear strategy is largely cal, nonfalsifiable, and data-free, both hawks and doves would dvised to question their own preexisting beliefs and policy pref- Cognitive maps can be useful guides to a bewildering world, but also get you lost.

fn. 63), 93.
ıotes four assumptions that "make up the bedrock of what must pass for a body ge": the prenuclear world is beyond retrieval; nuclear weapons must be fitted into coalition military strategy; investments in command and control improvements vhile; and "the nuclear arsenals of the superpowers cannot be regarded as instru-ilitary decision"; Gray (fn. 63), 311–13.
,15.

CRISIS AND DEMOCRACY IN LATIN AMERICA

By PETER H. SMITH

Khosrow Fatemi, ed. *U.S.-Mexican Economic Relations: Prospects and Problems*. New York: Praeger, 1988, 210 pp.

Abraham F. Lowenthal. *Partners in Conflict: The United States and Latin America*. Baltimore, Md.: Johns Hopkins University Press, 1987, 240 pp.

Lars Schoultz. *National Security and United States Policy toward Latin America*. Princeton: Princeton University Press, 1987, 365 pp.

Barbara Stallings and Robert Kaufman, eds. *Debt and Democracy in Latin America*. Boulder, Colo.: Westview Press, 1989, 232 pp.

Joseph Tulchin. *Argentina and the United States: A Conflicted Relationship*. Boston: Twayne Publishers, 1990, 172 pp.

Robert Wesson, ed. *Coping with the Latin American Debt*. New York: Praeger, 1988, 205 pp.

LATIN America defies the academic imagination. Countries of the region have been turning toward political democracy under the most improbable economic conditions. Not only do nations of Latin America seem to have survived the "debt crisis" of the 1980s, but from Brazil to Mexico and even Argentina they have begun to implement structural reforms, abandoning outdated forms of populism and entering a would-be era of export-led efficiency. As scholars were expressing skepticism about the meaning of elections,[1] citizens were giving credence and currency to the electoral process. Contradicting the notion that authoritarian rulers do not give up power voluntarily, General Pinochet in Chile and Daniel Ortega in Nicaragua were leaving office out of reluctant but obsequious respect for voter preferences. Leaders throughout the region have in the meantime moved toward a more amicable relationship with the United States and, at least in Washington's eyes, adopted a more "mature" and responsible posture toward the world community. Analysts have struggled to keep pace with all these changes. Events are moving faster than ideas.

What can explain these developments? It is the purpose of this essay

[1] Edward S. Herman and Frank Brodhead, *Demonstration Elections: U.S.-Staged Elections in the Dominican Republic, Vietnam, and El Salvador* (Boston: South End Press, 1984); and Terry Karl, "Imposing Consent? Electoralism vs. Democratization in El Salvador," in Paul W. Drake and Eduardo Silva, eds., *Elections and Democratization in Latin America, 1980–85* (La Jolla, Calif.: Center for Iberian and Latin American Studies/Center for U.S.-Mexican Studies, 1986), 9–36.

, explore three central questions. First, what accounts for the recent rocess of democratization in Latin America? How can this have occurred during an era of economic crisis? Second, what has been the impact of international factors? And of the role of the United States? And hird, what are the long-term prospects for democracy in Latin America? Will these polities be prepared to meet the challenges of the 1990s and beyond?

From this perspective any sample of recent books on U.S.–Latin American relations—including the selection under review—is bound to exhibit an archaic quality. Three of the volumes covered in this essay result from conferences held in 1987, a date that now seems far away and long ago. Providing commentary rather than original research, many contributors to these symposia seem to have been hurrying to keep up with ever-changing developments. On the whole, their efforts meet with indifferent success, a result that might appear to emphasize the growing insignificance of academic writing. Such an inference would be mistaken for at least two reasons, however. First, the decade of the 1980s was a crucial era of transition. It was a time when the political and economic systems of Latin America endured wrenching change, when leaders of almost all persuasions jettisoned formulas of the past and chose to search for a new future. Processes established during the 1980s will guide prospects for change throughout the 1990s and beyond. To comprehend the 1990s, then, we need to understand the 1980s.

Second, like the rumor of Mark Twain's demise, the conventional wisdom about Latin America is greatly exaggerated. Economic prospects are modest at best, even for the most promising of countries; they have been dealing not only with a debt crisis but also with a fundamental development crisis, and the path to recovery is not at all clear. Political transitions have been partial and limited. Social consensus is fragile and uncertain. Nations of Latin America continue to face profound and durable obstacles to development. A careful reading of these books provides a sobering perspective on the outlook for continuing change and growth throughout the region. The problems of the 1980s could endure into the next century.

But the most remarkable feature of the books under review, and of other recent publications in the Latin American field, is the absence of grand theory. This is an area that had for decades encouraged the application and generation of ambitious conceptual frameworks, from modernization theory in the 1950s and 1960s to *dependencia* theory in the 1960s and 1970s. By the mid-1980s both of these grand theories had gone out of fashion, and scholars came to display conspicuous caution. Apart

from modest attempts to test occasional hypotheses and propositions, there is not much theoretical content in the current literature.

This tendency could be a healthy sign, a recognition that events do not fit into preconceived frameworks. Certainly it gives the writings of the 1980s a tone of humility (as well as intermittent blandness). Yet this theoretical paucity also reflects a basic reality: we simply do not know what to think of Latin America—of its recent sociopolitical evolution or its evolving place in the international arena. What are the underpinnings of democracy in Latin America?

Models of Political Change

Putative explanations abound. Imbued with the optimistic hubris of the 1960s, students of Latin American politics, especially those from the United States, found ready and congenial answers in what had come to be known as modernization theory. As applied to Latin America by John J. Johnson and others, the argument posited simple causal connections. Economic development creates middle-class sectors that in turn espouse political democracy, either as a tactical means of gaining power or as an expression of enlightened values (the difference did not seem to matter at the time). The greater the level of economic development, the greater the likelihood of democratic politics. Indeed, the paradigm, possessed of internal coherence and logical structure, amounted to a full-blown theory of modernization. Moreover, it appeared to find empirical support in the cross-national analyses of Seymour Martin Lipset and Phillips Cutright, it carried implications for U.S. policy and foreign aid, and it offered bright hopes for the future.[2] It seemed to be too good to be true.

So it proved to be. Instead of dispensing prosperity, economic development (such as it was) accentuated the concentration of wealth and exacerbated existing inequalities. The middle strata, relatively privileged, forged little if any sense of class consciousness and, in critical moments of decision, joined with ruling classes in opposition to the popular masses. Political outcomes took a decidedly authoritarian turn, as shown by the lamentable experiences of Brazil (1964), Argentina (1966), and

[2] John J. Johnson, *Political Change in Latin America: The Emergence of the Middle Sectors* (Stanford, Calif.: Stanford University Press, 1958); Seymour Martin Lipset, *Political Man: The Social Bases of Politics* (Garden City, N.Y.: Doubleday, 1963), 27–63; Phillips Cutright, "National Political Development: Measurement and Analysis," *American Sociological Review* 28 (1963), 253–64. There were many variants on modernization theory, of course, and it is to be noted that Samuel P. Huntington placed as much emphasis on political institutionalization as on socioeconomic development; see Huntington, *Political Order in Changing Societies* (New Haven: Yale University Press, 1968).

Chile (1973). And in stark contradiction to modernization theory, these patterns emerged in the most developed—and most rapidly developing—countries of the continent. What could have gone wrong?

Two sets of answers came forth. One focused on the cultural traditions of Latin America and argued, in effect, that antidemocratic politics was (and remains) entirely consistent with a Catholic and Mediterranean world view that stresses the need for harmony, order, and the elimination of conflict. By failing to grasp these continuities, scholars had confused form with substance and rhetoric with reality. Latin America's constitutions were never as democratic as they appeared; party politics was not as representative as it might have looked. There had been no ownfall of democracy because there had been no upsurge in the first lace.[3] The academic community, afflicted by its own myopia and biases, ad simply misread the social facts.

Second was the *dependencia* approach, which accepted modernization ıeory's linkage of socioeconomic causes with political outcomes but ırned the answer upside down: since Latin America's economic develpment was qualitatively different from that of North America and Vestern Europe, it produced different results. Specifically, according to ıis argument, Latin America's experience was determined by the perasive fact of its economic "dependency." An inherent characteristic of ependent development is its tendency to intensify inequities, allocating enefits to sectors involved in the world market and denying them to ıarginal groups. As Guillermo O'Donnell asserted in the early 1970s, uling elites eventually come to face a clear-cut choice: sacrifice growth r pursue it by repressing the working classes (thus reducing wages, conrolling inflation, and attracting international investment). The prefernce, almost invariably, is for the latter course, even if it means violent ssaults upon already-organized working-class groups. Thus occurred he vicious coups and repressive regimes in Brazil, Argentina, and Chile. hey did not emerge in spite of Latin America's economic development; hey emerged because of it.[4]

By the late 1970s the social-science community had formulated a series f explanations for the turn toward authoritarianism in Latin America Two well-known anthologies neatly capture then-prevalent concerns.

[3] See Glen Dealy, "Prolegomena on the Spanish American Political Tradition," *Hispanic merican Historical Review* 48 (February 1968), 37–58; Howard J. Wiarda, "Toward a ramework for the Study of Political Change in the Iberic-Latin Tradition: The Corporative Aodel," *World Politics* 25 (January 1973), 206–35; and Wiarda, ed., *Politics and Social Change Latin America: The Distinct Tradition* (Amherst: University of Massachusetts Press, 1974).

[4] Guillermo O'Donnell, *Modernization and Bureaucratic-Authoritarianism: Studies in South merican Politics* (Berkeley: Institute of International Studies, University of California, 1973).

One focused on the breakdown of democracy; another dealt with the rise of the "new" authoritarianism.[5] Democratization in the 1980s took the field by surprise.

In Search of Economic Determinants

Economic conditions have hardly offered propitious circumstances for the emergence of democracy. Over the last decade Latin America has faced a two-part crisis. One stems from the exhaustion of the strategy of import-substitution industrialization (ISI). Ever since the depression, and especially after World War II, major countries of the region sought to promote economic growth by stimulating domestic demand and protecting local industry. This strategy required an active, interventionist state and often found ideological justification in populism and/or nationalism. In many ways ISI achieved considerable success, especially in the 1950s and 1960s. Average annual rates of GDP growth for the region climbed steadily from just under 5 percent in the late 1940s to more than 6.5 percent for the early 1970s. From 1965 to 1980 Mexico enjoyed an annual average growth rate of 6.5 percent, Brazil had 9.0 percent, Argentina had 3.5 percent, and even Peru—today's quintessential basket case—stood at 3.9 percent. But by the late 1970s and the 1980s ISI could no longer provide keys to growth: domestic markets were saturated, industries were inefficient, budgets were awash in deficit.[6]

Enter the debt crisis. Eager to recycle petrodollars, major banks found willing borrowers in Latin America. Between 1970 and 1980 the region increased its external debt from $27 billion to $231 billion, with debt-service payments of $18 billion. Complications ensued: real interest rates rose, commodity prices declined, capital dried up.[7] Caught in this squeeze, Latin America experienced increasing difficulty in meeting obligations on the debt. In August 1982 Mexico proclaimed its inability to pay. Since then the region has been in the grips of a debt crisis that has, in turn, greatly exacerbated and intensified the long-term development crisis.

While public officials now express considerable optimism about long-

[5] Juan J. Linz and Alfred Stepan, eds., *The Breakdown of Democratic Regimes* (Baltimore, Md.: Johns Hopkins University Press, 1979); David Collier, ed., *The New Authoritarianism in Latin America* (Princeton: Princeton University Press, 1979).

[6] James W. Wilkie and Enrique Ochoa, eds., *Statistical Abstract of Latin America*, vol. 27 (Los Angeles: UCLA Latin American Center, 1989), 939, fig. 33.2; and World Bank, *World Development Report 1989* (New York: Oxford University Press, 1989), 166–67, table 2.

[7] See Pedro-Pablo Kuczynski, *Latin American Debt* (Baltimore, Md.: Johns Hopkins University Press, 1988).

un prospects for Latin America's economic development, scholarly re-
lection exudes caution and concern. Robert Wesson's *Coping with the
atin American Debt*, a collection of papers from a conference in late
987, offers compelling testimony to the unresponsiveness and unwork-
bility of the successive debt programs, plans, policies, and postpone-
nents of the 1980s. It also provides vivid examples of the polarization
nd frustration that came to characterize both sides of the negotiating
able. Listen, for example, to Carlos E. Melgar, Peruvian senator: "Pov-
rty and backwardness are, to a great extent, the result of the exploitation
nd spoliation to which they have been subjected by developed countries,
nd which today finds in the foreign debt its most sophisticated and in-
uman expression" (p. 151). And to Charles J. L. T. Kovacs, American
anker, who contended that "it is not too much of an exaggeration to
uggest that without capital flight, we would not have a serious debt
roblem today" (p. 42).

On balance, the contributions convincingly demonstrate that the mud-
lling-through strategies of the 1980s, including the much-touted Baker
'lan, served the interests of creditors rather than of borrowers. "By 1987
he strategy was fairly well exhausted," Wesson writes. But because all
arties have an interest in keeping the negotiations going, they "will
loubtless continue their minuet of negotiations, conditions, and accords
o keep up appearances without doing anything more than turning in
ircles" (pp. 186, 188). And so indeed they have, up to and through the
o-called Brady Plan.

We can now assess the results. Latin America's total external debt
limbed from $231 billion in 1980 to approximately $417.5 billion in
990. External debt as a proportion of exports increased from 209.8 per-
ent to 272.4 percent. Debt service accounted for 34 percent of exports in
980 and 42 percent in 1990—after a decade of export promotion. From
982 through the end of the decade Latin America transferred over $200
illion to industrialized nations, equivalent in constant dollars to several
imes the Marshall Plan.[8] Real GDP per capita declined in 1981, 1982, 1983,
988, and 1989—and showed a cumulative decline of 5–10 percent over
he course of the decade.[9] This can hardly qualify as progress.

[8] International Monetary Fund, *World Economic Outlook, October 1989* (Washington, D.C.: nternational Monetary Fund, 1989); and Inter-American Dialogue, *The Americas in 1989: onsensus for Action* (Washington, D.C.: Aspen Institute, 1989), esp. 3–4.

[9] Figures vary slightly by source. See Wilkie and Ochoa (fn. 6), 991, table 3363; Comisión conómica para América Latina (CEPAL), *Notas sobre la economía y el desarrollo*, 485–86, *alance preliminar de la economía de América Latina y el Caribe, 1989* (December 1989), cuadro ; and Inter-American Development Bank, *Economic and Social Progress in Latin America, 989 Report* (Washington, D.C.: Inter-American Development Bank, 1989), 463, Table B-1.

In this saga, of course, Mexico holds a special place. It was Mexico that unleashed the debt crisis in August 1982, when Finance Minister Jesús Silva Herzog made his famous trip to Washington, and it is Mexico that has become the first major debtor to meet the provisions of the Brady Plan. Resulting from a conference at Laredo State University in mid-1987, the essays in the volume edited by Khosrow Fatemi explore the complexity and variety of U.S.-Mexican economic relations as they pertain to trade, investment, immigration, oil, debt, and border matters. Other than its topical exploration of the process of "invisible integration" between the two economies, the collection lacks explicit conceptual unity. The contributions imply, however, that Mexico and the U.S. are so closely interconnected that no single issue, not even the debt, could lead to a rupture or even a serious downturn in deeply embedded patterns of interdependence.

On the whole, *U.S.-Mexican Economic Relations* presents an optimistic assessment of its subject. According to Robert Looney, a reasonable package of reforms together with productive use of new capital in Mexico "should enable the country to achieve a long-run solution to its debt problems, while at the same time assuring a sustainable flow of external capital" (p. 33). Other contributors maintain that migration will continue, to the benefit of both countries, that *maquiladoras* (in-bond assembly operations) will expand, especially in the interior of Mexico, and that technology transfer will accelerate, to and through Mexican as well as foreign-owned firms.

As the case of Mexico demonstrates, even under Carlos Salinas Gortari, there are serious problems ahead. Perhaps the most critical ch lenge will be to identify a sustainable niche within the evolving inter tional economy. Mexico might be able to carve out a special position an export platform to the American market, especially through the tential realization of a free-trade agreement with the U.S. (the possibil of which is nowhere mentioned in this volume).[10] Despite the vision o free-trade zone for the entire Western Hemisphere, proclaimed in mi 1990 by President George Bush, the rest of Latin America might r have such sanguine options. What role might there be in the wake of t Asian NIEs?[11] What if venture capital heads for Eastern Europe inste of Latin America?

[10] For comparison, see William Glade and Cassio Luiselli, eds., *The Economics of Inten pendence: Mexico and the United States*, vol. 2 of Rosario Green and Peter H. Smith, ec *Dimensions of United States-Mexican Relations*, 5 vols. (La Jolla: Center for U.S.-Mexic Studies, University of California, San Diego, 1989).

[11] See Robin Broad and John Cavanagh, "No More NICs," *Foreign Policy* 72 (Fall 198 81–103.

A second basic challenge concerns social justice. The debt-coping strategies of the 1980s have inflicted painful sacrifice upon the people of Latin America, especially the working classes. Unemployment and underemployment are at record levels. Real wages have declined precipitously—by as much as 50 percent in some countries. This is a reality that policymakers in Washington and elsewhere would prefer to ignore. But as Riordan Roett wrote in 1987 of policy options on the Brazilian debt: "The choice seems to be between order or ethics" (Wesson, 81). The Brady Plan notwithstanding, policymakers still face that same choice, with reference to almost all of Latin America.[12]

If economic conditions are staying so bad, why are political conditions getting so good? In the late 1970s, it is often said, approximately 90 percent of the people of Latin America lived and labored under authoritarian regimes; by the mid-1980s nearly 90 percent enjoyed some sort of democracy, and by 1990 the figure was higher than that. The transitions took distinctly different forms, from the embarrassed retreat of the Argentine generals to the orchestrated exit of their Brazilian comrades to the electoral upheavals in Chile and Nicaragua. How could democratization take place amid such economic circumstances?

One possibility is that the debt-cum-development crisis presented authoritarian rulers with an unsolvable problem that hastened their exit from office. Military leaders, in particular, had seized power with the avowed intention of asserting law and order, extirpating subversives, and restoring economic growth. They managed to impose order but they flouted law, disappeared opponents, and achieved mixed results in economics. The Brazilian generals oversaw a period of phenomenal growth in the late 1960s and early 1970s; the Pinochet regime appears to have brought about a significant degree of restructuring. But even these accomplishments entailed excessive human costs for the citizenry and high political costs for the regimes. In the end, it could be said, the severity of Latin America's economic predicament determined and accelerated the demise of authoritarian regimes—and thus paved the way for democracy.

Economic explanations for the *consolidation* of the new democracies have taken alternative forms. Perhaps most surprising has been a reassertion of modernization theory, or what might be called neo-modernization theory. "According to a large body of comparative research on democracy," Mitchell Seligson has written,

[12] One thoughtful effort to confront this problem appears in Economic Commission for Latin America and the Caribbean (ECLAC), *Changing Production Patterns with Social Equity* (Santiago de Chile: United Nations, Economic Commission for Latin America and the Caribbean, 1990).

> until relatively recently the minimum levels of socioeconomic developme that are necessary to sustain democratic rule simply were not present most of Latin America. Several studies have led to the near-universal co clusion that underdevelopment and stable democracy do not mix. The appears to be a lower threshold of economic and sociocultural developme beneath which stable democratic rule is unlikely to emerge. The incon threshold appears to be around $250 per capita in 1957 dollars, and t sociocultural threshold, usually defined by educational achievement, seen to be the reduction of illiteracy to below 50 percent.

By the 1980s, he continues, "the socioeconomic foundations for st democracy had finally been established" in Latin America. This does "lead to the conclusion that democratic rule is somehow inevit among these nations. ... [S]ocioeconomic development is only a ne sary, not both a necessary *and* sufficient condition for the developn and maintenance of stable democracy."[13]

Like classical modernization theory, this argument suffers from unfortunate failure to distinguish between (1) the demise of authorit anism, (2) the transition toward democracy, and (3) the consolidatio democracy itself. Levels of socioeconomic development are likely t most relevant for successful consolidation (3) and least relevant to transitional process (2). Nonetheless, this line of reasoning tacitly of retroactive vindication for modernization theory. Its premises were curate, its application premature. It did not suit the 1960s, but it fits 1980s.

A contrasting position holds that Latin America's economic crisis exert a critical impact on the consolidation (or not) of democracy. Ind some argue that the region's fragile democracies are especially susce ble to pressures and problems deriving from the foreign debt—a more broadly, from the contemporary crisis of economic developm Democracy cannot sustain the debt; the debt cannot sustain democr Such reasoning stresses the idea that democracy depends on popular s port, that austerity programs antagonize the masses, and that aliena masses might withdraw support from the democratic regime. This co lead to popular uprisings and/or military interventions. Yet there alternative outcomes:

> 1. The armed forces, not widely renowned for the success of their economic policy, might be extremely reluctant to assume responsibility fo management of the debt.
> 2. Democratic leaders, in search of popular support, might seek to rally

[13] Seligson, "Democratization in Latin America: The Current Cycle," in James M. Mal and Mitchell A. Seligson, eds., *Authoritarians and Democrats: Regime Transition in Latin Am ica* (Pittsburgh, Pa.: University of Pittsburgh Press, 1987), 7–9.

nationalist sentiment by condemning the role of foreign banks and international agencies (as Alan García attempted in Peru).

3. Disaffected masses may not necessarily create a crisis for the system. Rather, they might just vote out the party in power.

It is worth noting, in fact, that major elections in 1989 and 1990 led to widespread triumph of opposition movements in Argentina, Brazil, Chile, Peru, and Uruguay. Even official results from the Mexican presidential elections of 1988 showed a precipitous reduction in the ruling party's popular base.

In the short term the debt burden does not automatically create a structural crisis for the new democracies. In the medium run, perhaps, a democratic system might be better equipped than an authoritarian regime to respond to popular dissatisfaction. And in the long run the sociopolitical costs of continuing debt service could undermine the legitimacy of *any* kind of system, democratic or authoritarian.

We might therefore expect to see two broad patterns of political behavior in the 1990s and beyond. First, with the Latin democracies, there might be a fairly constant rotation of parties in power, as opposition parties succeed in tarnishing ruling parties with the social costs of the debt. Such enfranchisement of the opposition and alternation in power could actually assist the consolidation of democratic regimes. The economic crisis might foster tolerance and accommodation while imbuing political rivals with a shared sense of national purpose. In some cases the Latin American debt crisis may thus actually strengthen democratic processes and systems.

In the longer term we might witness a fairly volatile alternation of authoritarian and democratic regimes, as neither will be able to "solve" the debt problem or the economic crisis. Such a pattern would reaffirm the historically cyclical pattern of political change in Latin America, from authoritarianism to democracy and back again, but perhaps with some added twists. Because of economic crisis and low system legitimacy, the cycles might well become shorter—that is, the changes more rapid—and because of socioeconomic mobilization, popular participation might become greater. This could enhance demands for democratization. It could also lead to greater violence.

The Turn toward Political Explanations

In the effort to account for democratization, many analysts have turned away from economic (or economistic) arguments in favor of political fac-

tors. In one broad and ambitious analysis, Guillermo O'Donnell and Pl
lippe Schmitter focus not on structure but on statecraft. Drawing esp
cially on the experiences of Argentina and Brazil, they assert th
political entrepreneurship acquires a key role in the transitional proce
As "the question of democracy" returns to the national agenda, eli
initiate transitions. Intramilitary resistance to hard-liners leads soft-line
to search for allies among moderate elements of the civilian elite. Th
begins a complicated minuet of political maneuvering, where both sid
risk "coup poker" and attempt to increase their political space. Popul
mobilization does not precipitate this process but eventually becomes
important element within it, as intellectuals and artists provoke "the re
urrection of civil society" that itself constitutes a new power resourc
Elections and the return of party politics conclude the process of trans
tion; the greater the representation for the forces of authoritarianism, th
more likely the prospects for successful consolidation. Transitions th
occur as a result of ideological aspiration and political leadership, n
because of economic contradictions. (Since political and social process
"are neither symmetric nor reversible," they explain, the causes of a dem
ocratic breakdown may not necessarily be the same as the reasons for
transition from authoritarianism.)[14]

O'Donnell and Schmitter emphasize bargaining, negotiation, an
compromise. Utilizing the metaphor of a multilayered chess game, the
stipulate that two moves are forbidden. First, it is illegal to take, or eve
checkmate, the king of one of the players: "In other words, during th
transition, the property rights of the bourgeoisie are inviolable." Second
it is illegal to take or even circumscribe the queen: "In other words, t
the extent that the armed forces serve as a prime protector of the right
and privileges covered by the first restriction, their institutional existence
assets, and hierarchy cannot be eliminated or seriously threatened."[1]
These are serious restrictions. In view of the horrific experience of the
bureaucratic-authoritarian regimes, however, the restoration of civil lib-
erties is for O'Donnell and Schmitter well worth its high price.

One conspicuous feature of this literature is its preoccupation with the
dynamics of the transition itself, as distinct from the breakdown of au-
thoritarianism or the consolidation of democracy. This choice has some-
thing of a self-fulfilling quality, since political entrepreneurship is more

[14] O'Donnell and Schmitter, "Tentative Conclusions about Uncertain Democracies," in
O'Donnell, Schmitter, and Laurence Whitehead, eds., *Transitions from Authoritarian Rule*
Prospects for Democracy (Baltimore, Md.: Johns Hopkins University Press, 1986), pt. IV, p.
18. See also O'Donnell, "Tensions in the Bureaucratic-Authoritarian State and the Question
of Democracy," in Collier (fn. 5), 285–318.

[15] O'Donnell and Schmitter (fn. 14), pt. IV, p. 69.

likely to play a determinative role during the transition than during the other phases.[16] But as Alfred Stepan points out, the focus derives its justification from the premise that "the actual route" of democratization can exert "independent weight" on the shape of the final outcome. The path of democratization can affect the eventual role of the military, the scope of permissible opposition, and the composition of the ruling coalition. In other words, the transition itself constitutes an independent variable in the determination of the ensuing democratic system.[17]

This train of reasoning has produced a number of provocative hypotheses. My own distillation of contemporary literature yields a range of examples:

H_1. The harsher the authoritarian regime, the greater the degree of subsequent democracy.

H_2. The greater the authoritarian repression, the greater the discontinuity between pre- and postauthoritarian democracies.[18]

H_3. The longer an authoritarian regime remains in power, the greater the transformation of the preauthoritarian civilian political system.

H_4. The more sudden and unexpected the transition from authoritarianism, the greater the likelihood of a popular upsurge—and of its impact on the subsequent democracy, where it would be likely to support prodemocracy purists instead of moderate forces.

H_5. The more controlled the transition from authoritarianism, the easier the installation and consolidation of democracy.

H_6. The more total and sudden the collapse of the authoritarian regime, the more likely the appearance of a full and open democracy—subject to severe authoritarian reversals (either because of the absence of intraelite "pacts" for moderation or because of unrealistic expectations for the new democracy).

H_7. The more unsuccessful the authoritarian episode, the less the resistance to democracy—and the less likely a subsequent reversal.

H_8. The more radical the economic policies of the authoritarian regime, the greater the reluctance to initiate or accept a process of transition.

In a similar vein, Enrique Baloyra has called for "a predominantly olitical explanation" of transitions toward democracy ("Nothing else vill do"). Focusing upon what he calls the "endgame," the final and lecisive confrontation between reformers and obstructionists, Baloyra tresses the long-term significance of transition itself:

[16] On this and related points, see Daniel H. Levine, "Paradigm Lost: Dependence to Denocracy," *World Politics* 40 (April 1988), 377–94.

[17] Stepan, "Paths toward Redemocratization: Theoretical and Comparative Considerations," in O'Donnell, Schmitter, and Whitehead (fn. 14), pt. III, p. 65.

[18] On the survival of political parties, see Marcelo Cavarozzi and Manuel Antonio Garretón, eds., *Muerte y resurrección: los partidos políticos en el autoritarismo y las transiciones del Cono Sur* (Santiago de Chile: Facultad Latinoamericana de Ciencias Sociales, 1989).

> The exercise will force us to revise some currently held beliefs about the inadequacy of democracy for the politics of late, dependent capitalism. If a logic of transition could prevail under such adverse circumstance, there is reason to believe that these weak inaugurations could be consolidated if *the lessons of transitions were assimilated*. . . .
>
> In essence, a decision to move away from a regime of exception in a *capitalist state that does not imply a profound change in the political economy is bound to be linked to a strategy of extrication through elections.* . . . [Transition] vindicates the lesson that there are no truly viable alternatives to civilized politics. This is the most important lesson of the logic of transition.[19]

In effect, the transition can establish and implement rules of the gam that can provide the foundation for durable consolidation of democrac

Yet another kind of emphasis on political factors, dealing especiall with prospects for consolidation, tends to focus on institutional variable Here the debate centers on the relative merits of presidentialism, sem presidentialism, and parliamentarism. Juan Linz, Arend Lijphart, an Arturo Valenzuela have argued forcefully in favor of the parliamenta alternative.[20] Constitutional debates in Brazil and especially in Argentin have focused on varieties of semipresidentialism.[21] The point is that suc institutional arrangements could promote power sharing, encourage ac commodation, and thus enhance the prospects for durable consolidatio The irony, and perhaps the tragedy, is that genuine parliamentarism ha yet to be attempted in Latin America.[22]

In response to many such concerns, from economic to political argu ments, Barbara Stallings and Robert Kaufman brought together a prom inent group of analysts in mid-1987 to deal with two basic issues: (1) th influence of regime type on policy choices for dealing with the debt crisis

[19] Baloyra, ed., *Comparing New Democracies: Transition and Consolidation in Mediterranea Europe and the Southern Cone* (Boulder, Colo.: Westview Press, 1987), 3, 18, 47.

[20] Oscar Godoy Arcaya, ed., *Hacia una democracia moderna: la opción parlamentaria* (Santiago: Universidad Católica de Chile, 1990); Linz, "Democracy, Presidential or Parliamentary: Does It Make a Difference?" (Paper presented to ACLS conference on Constitutions and Constitutionalism in Latin America, Punta del Este, Uruguay, November 1988); idem "The Perils of Presidentialism," *Journal of Democracy* 1 (Winter 1990), 51–69; Larry Diamond and Linz, "Introduction," in Diamond, Linz and Seymour Martin Lipset, eds., *Democracy in Developing Countries*, vol. 4, *Latin America* (Boulder, Colo.: Lynne Rienner, 1989), 24–27; and Lijphart, "Basic Choices in Democratic Constitutional Design," *Journal of Democracy* (forthcoming).

[21] Carlos Santiago Nino, "Transition to Democracy, Corporatism and Constitutional Reform in Latin America," *University of Miami Law Review* 44 (September 1989), 129–64; *Reforma constitucional: Dictamen preliminar del Consejo para la Consolidación de la Democracia* (Buenos Aires: Eudeba, 1986); and *Reforma constitucional: Segundo dictamen del Consejo para la Consolidación de la Democracia* (Buenos Aires: Eudeba, 1987).

[22] For a view that emphasizes institutional factors in historical perspective, see Myron Weiner, "Empirical Democratic Theory," in Weiner and Ergun Ozbudun, eds., *Competitive Elections in Developing Countries* (Durham, N.C.: Duke University Press, 1987), 3–34.

and (2) the impact of the debt crisis on regime type and political transition. To their credit, and unlike many others, Stallings and Kaufman offer at the outset their working definition of democracy: "A regime can be considered democratic to the extent that incumbent governments must win and retain power through competitive elections, tolerate opposition challenges to their incumbency, and deal with relatively independent interest groups" (p. 1). Their recognition of the need for definitional precision is entirely laudable in principle, if less satisfactory in practice. Their concept tends to focus on electoral formalities while ignoring the importance of starting points and of trajectories.[23] And the notion of "transitional democracy," which they introduce in their concluding essay, suffers from teleological presumptions. Transition to what? How do we know?

It is essential to identify the structural properties of both the transitions and the resulting democracies. Latin American politics of the 1980s has displayed three outstanding characteristics. One is the rise in civic participation, from the compelling protests of the Madres de la Plaza de Mayo in Argentina to the understated agony of members of the working class in Chile to the long and dignified lines of voters in Central America. Second is the increasing commitment to the electoral process. Elections can serve many functions, of course, from the selection of leaders and programs to the legitimation of behind-the-scenes influence. But the intensity of electoral campaigns and of voter participation throughout the 1980s, often conducted under hazardous conditions, indicates a remarkable willingness on the part of ordinary citizens to utilize ballots to express preferences, settle conflicts, and confer authority on leaders.[24] Third is the withdrawal of military leaders from formal office and, as a result, the emergence of civilian presidents. This has occurred most clearly in Peru, Guatemala, Brazil, Argentina, and Chile. It has brought middle-class civilians to power, and it has provoked a conspicuous change in the social composition of political leadership.

But these processes neither comprise nor guarantee full democracy. There are at least five major limitations on contemporary democracy in Latin America. First, paradoxically enough, is the continuing role of the military. Almost everywhere in Latin America the armed forces have established a broad range of fiefdoms and prerogatives above and beyond

[23] See Stepan (fn. 17), 64–84.

[24] See John A. Booth and Mitchell Seligson, *Political Participation in Latin America*, 2 vols. (New York: Holmes and Meier, 1978–79); and Enrique C. Ochoa, "The Rapid Expansion of Voter Participation in Latin America: Presidential Elections, 1845–1986," in James W. Wilkie and David Lorey, eds., *Statistical Abstract of Latin America*, vol. 25 (Los Angeles: UCLA Latin American Center Publications, 1987), 862–910.

the scope of civilian authority.[25] This is perhaps most apparent in Argentina, where throughout the 1980s the once-disgraced military staged a series of uprisings designed to restore the traditional boundaries of military power, though not to overthrow the regime; with the declaration of amnesty by beleaguered President Carlos Saúl Menem, the officers won their case. From Chile to Brazil to Guatemala and El Salvador, generals no longer sport presidential sashes but continue to wield enormous influence and autonomy.

Second, the recently installed democracies rest upon a tenuous foundation: the emasculation of the Left. In country after country (except Nicaragua), elections have entailed choices between the Center and the Right (and combinations thereof). The Left has been most conspicuous by its absence. During the authoritarian episodes of the 1960s and 1970s the Left was subjected to torture, intimidation, and in some cases near extermination. As a result, electoral contests have been tepid, fail-safe affairs. Subsequent events in Eastern Europe and the Soviet Union have eroded the ideological foundations of Marxism, which may give way to various forms of populism in the 1990s. In the meantime, there are no more Allendes.

Third has been tacit censorship. One of the most crucial problems of the day, redistribution of income—on behalf of either workers or peasants, either as revised wages or as land reform—is simply not open to public debate. Nor, in most countries, is there a possibility of legal prosecution of military murderers. (This helps explain why conservative generals are willing to stay in the barracks.) This is not to say that debt management, as sketched out by Stallings and Kaufman, is an unimportant matter. It it merely to say that the most crucial consequence of the debt/development crisis—its regressive impact on income distribution—is not a subject for meaningful public discussion.

Fourth, there continues to be widespread and semiofficial abuse of human rights. This is especially true in Central America, most especially in El Salvador and in Guatemala, but it is also evident in Colombia and Peru (where the U.S. is ardently supporting paramilitary campaigns as part of a noble crusade against international drug trafficking).[26] In many

[25] See Alfred Stepan, *Rethinking Military Politics: Brazil and the Southern Cone* (Princeton: Princeton University Press, 1988), esp. chaps. 6 and 7.

[26] See, e.g., Americas Watch Committee, *Tolerating Abuses: Violations of Human Rights in Peru* (New York: Americas Watch, 1988); idem, *The Killings in Colombia* (New York: Americas Watch, 1989); idem, *The Drug War in Colombia: The Neglected Tragedy of Political Violence* (New York: Americas Watch, 1990); Amnesty International U.S.A., *Guatemala: Human Rights Violations under the Civilian Government* (New York: Amnesty International, 1989); and Peter H. Smith, "Drug Wars in Latin America," *Iberoamericana* (Sophia University, Tokyo) 12 (Summer 1990), 1–16.

countries, Chile and Argentina now excepted, elected civilian presidents are either unwilling or unable to protect those citizens' rights that are most basic to the achievement and perpetuation of genuine democracy.

Fifth, these newly installed governments have revealed a marked propensity for undemocratic decision making, especially in regard to economic policy. From the technocrats in Mexico to Fernando Collor de Melo in Brazil to Alberto Fujimori in Peru, policymakers have ignored or flouted the popular will. "What we have really been seeing in Latin America over the decade of the 1980s," according to James Malloy, "has been the emergence of new kinds of hybrid regimes which in the short term at least marry an outward electoral democracy with a technocratic and authoritarian policy process especially in economic matters." Beset by problems of governance and the exigencies of neo-liberal austerity programs, civilian rulers display an "inner authoritarian bent."[27] In so doing, they jeopardize one of the key precepts of democracy, the idea of accountability.

These caveats convey at least three messages. First, they argue in favor of strict, rigorous, and consistent usage of such terms as *democracy* and *democratization*. Second, they reveal the recent democratic transitions to be contingent, uncertain, conditional phenomena—not automatic processes or self-fulfilling prophecies. Third, and in turn, they call for caution against sweeping generalization and for judicious use of comparative analysis. Whatever might be the "winds of change" in Eastern Europe, for example, it is not entirely self-evident that these same breezes are blowing in Latin America.

Finally, the restrictive nature of these democracies suggests that they might move in the direction of more complete democratization—or they might not. My own guess is that they are just as likely to transform themselves into *democraduras*, as O'Donnell and Schmitter have called them, civilian governments under the auspices of military authorities who are dedicated to the imposition and maintenance of social discipline. The possibility will grow stronger if the debt/development crisis continues to fester. Kaufman and Stallings offer a similar judgment: "The greater danger to democracy in the next decade is not military overthrow, but military subversion" (p. 219). There might well emerge a variety of "hard-line democracies," quasi-democratic and quasi-authoritarian, as occurred in Uruguay under José María Bordaberry some time ago—electoral regimes that make short shrift of human rights and dem-

[27] James M. Malloy, "Economic Crisis and Democratization: Latin America in the 1980s," *The Latin American and Caribbean Contemporary Record* 8 (forthcoming).

ocratic freedoms. In other words, the future of Latin America might not be Costa Rica; it might well be Guatemala or Peru.

The most original essays in the Stallings-Kaufman volume offer provocative insight into the distributional impacts of the debt crisis and the muddling-through strategies of the 1980s. Jeffry A. Frieden argues that winners are those who can easily move assets, who take their currency or goods to Miami and exit from the political game (which explains capital flight from Mexico and Venezuela). Losers have geographically fixed assets (as shown by relatively low capital flight from Chile and Brazil), and, as a result, they stay on to engage in political contests. The subsequent question is whether they will try to work from within the regime, as in Chile, or join the opposition, as in Brazil. Sylvia Maxfield supports this general argument, indicating that debt-led growth in Latin America was especially favorable to the financial sector and to large firms. And Ian Roxborough demonstrates convincingly that organized labor has been one of the biggest losers: in Mexico, for instance, real wages plummeted by nearly 40 percent between 1982 and 1985. These papers thus trace the outlines of a sociopolitical interpretation of debt-related problems of the 1980s. Unfortunately, there is no essay on that most voluble and vocal of sectors, the urban middle class. Nor is there any explanation of why Latin American governments, democratic and authoritarian, chose to subject their citizens to such sacrifice and deprivation, or why these selfsame regimes made such timid and ineffective efforts to join together in a debtors' cartel.

Debt and Democracy raises two broad, analytical issues. One is the question of causality. The debt crisis (and/or the development crisis) may or may not "cause" democracy or dictatorship in Latin America. As Paul Drake demonstrates through his reexamination of the 1930s, debt crises can jeopardize either democratic or authoritarian regimes. But more to the point: the absence of military coups in the 1980s does not mean that the distributional effects of the debt crisis do not foster confrontation, polarization, and instability in the political arena; it only means that there are countervailing factors. In time, Riordan Roett's prediction could still come to pass:

> Given an option, voters may favor a populist leader who promises a better future. The issue is less whether democratic regimes will be forcibly overthrown than whether Latin America will slowly return to the irresponsible, populist politics of a bygone era. Promises of distribution cannot be met given the current economic crisis. What will follow the failure of populist policies? Will populist regimes so radicalize the policy process that

polarization will inevitably result, with the armed forces becoming a real threat in the 1990s? (pp. 70–71)

Long-run prospects are not necessarily good. And in the short term the debt crisis still presents a serious threat to democratic consolidation.

A second question concerns the autonomy of the political arena. Noting the persistence of "democracy" amidst the ravages of debt and muddling through, Kaufman and Stallings surmise that "there may turn out to be more separation between the economic and political spheres than is generally believed" (p. 217). Of course this all depends on what is generally believed by whom. Their argument is nonetheless consistent with recent literature on the autonomy of the state and on the desirability of "bringing the state back in."[28] But that does not argue in favor of a total separation of political and economic spheres, which is not justifiable on either empirical or logical grounds. What it really suggests is the need to ascertain interconnections with clarity and precision.

Focus on the International Arena

Literature on democratization in Latin America tends to concentrate on ' e interplay between state and opposition within the domestic arena. onically enough, for a field so recently dominated by the *dependencia* proach, there has been conspicuously little attention given to the power d presence of international political factors. (Even *Debt and Democ-cy*, with its emphasis on external pressures of the debt crisis, deals ainly with responses and realignments of forces within national poli-s.) But international and contextual variables could have a major im-ct on political transformations throughout the region, and fulsome clarations from Washington assert that the U.S. has played a major d positive role in the promotion of democracy throughout the region.[29] ooks under review dispute this latter contention.

As Lars Schoultz shows in his study *National Security and U.S. Policy ward Latin America*, the idea of security—rather than aid, assistance, partnership—has long dominated U.S. thinking toward the region. Having written a book on human rights, he came to regard that as "a sidual category.... If one wants to study the core of United States olicy toward Latin America, one studies security" [p. xi].)[30] His analysis

[28] See Peter B. Evans, Dietrich Rueschemeyer, and Theda Skocpol, eds., *Bringing the State ck In* (Cambridge: Cambridge University Press, 1985).

[29] For an insightful overview and critique, see Cole Blasier, "The United States and De-ocracy in Latin America," in Malloy and Seligson (fn. 13), 219–33.

[30] See Schoultz, *Human Rights and United States Policy toward Latin America* (Princeton: rinceton University Press, 1981).

traces both the *content* and the *transformation* of the Washington consensus that depicted international communism as both cause and consequence of political instability in Latin America. On the basis of interviews, documents, and firsthand observation, Schoultz writes of the immediate postwar period: "Virtually all policy makers shared two core beliefs: they believed that instability in Latin America was caused by Communist adventurism and, as a result of this, they believed that instability, left unchecked, was a threat to U.S. security" (p. 14). Over time, especially as nongovernmental groups joined in public debates, two key questions emerged within the U.S. policy-making establishment:

1. whether political instability in Latin America resulted from communist adventurism or from structural poverty; and
2. whether political instability in Latin America would lead to communist advances (bad for the U.S.) or to reform and structural change (which might be good for the U.S.).

Arraying groups along these two dimensions, Schoultz distinguishes "conservatives" (who regard communism as cause and consequence of instability) from "moderates" (who regard poverty as the cause but communism as the possible effect) from "liberals" (who see poverty as the cause and reform as the result). "This group is new," Schoultz contends of the liberals, "and if it persists, it threatens to change forever the face of U.S. policy toward Latin America" (p. 21).

Schoultz explores the potential impact of Latin American "instability on U.S. "security" with surgical precision. He wields his scalpel on ideas about the need for the United States to maintain strategic access to key commodities:

—*Raw materials*. "The data indicate that U.S. dependence upon Latin America for essential strategic raw materials is fairly modest, and Latin America's supplies are extremely highly concentrated in Brazil, Jamaica, Mexico, and Venezuela" (p. 157).

—*Military bases and military support*. Despite the presence of bases in Panama, Honduras, and Guantánamo, "a rough consensus has been reached [within the U.S. military!] that there is little need for U.S. military bases in Latin America" (p. 173) and that Latin American armed forces would be neither effective nor reliable in large-scale combat.

—*Vital maritime routes*. Sea lines in the Caribbean (but not elsewhere) are important, as they would be in wartime, but they are already well defended by U.S. forces.

In addition to insisting on strategic access (for the U.S.), security doctrine has stressed the importance of strategic denial (for the USSR). The Soviets would in fact have little to gain from the installation of new

intelligence facilities or military bases, though there might be some advantage in the deployment of decoys that could tie down U.S. forces in the Western Hemisphere. Nor is it likely that a U.S. "loss" in, say, Central America would tip a fragile strategic balance of power in favor of the Soviets. Despite its improbable foundations, the logic of strategic denial has exerted a profound impact on U.S. policy. To quote one of Schoultz's more memorable passages:

> Most policy makers believe that the security of the United States would not be adversely affected if Latin America were, say, to sink to the bottom of the sea.... Why? Because if the region were to disappear, then there would no longer be a need to worry about Latin America's vulnerability to Soviet and Soviet/Cuban adventurism. And that is the true significance of Latin America to U.S. security: it is not that policy makers want much of anything from Latin America but rather that they do not want the Soviet Union to have it. (p. 310)

This whole debate now seems hopelessly anachronistic, the relic of a cold war since dumped into history's ash can. But Schoultz's masterful exposition of the Washington mind-set commands attention for several reasons. First, it underscores the intensity (and shortsightedness) of U.S. concern with security issues. Second, it reveals the paucity of U.S. concern with the promotion of democracy. One is reminded of John F. Kennedy's stated willingness to tolerate a Trujillo if that would prevent another Castro. Third, it draws attention to the role of the U.S. in the contemporary process of democratization.

In fact the influence of international factors—and of the U.S.—might be more evident with regard to what did not happen than to what actually occurred. One of the most notable features of the recent transitions is a nonoccurrence. The United States did not intervene to prevent, control, or truncate the process of democratization as it has done on so many other occasions. This self-restraint came at a time when the United States was stepping up its activity in Central America and mounting an invasion of Grenada and when prevailing doctrinal formulations in Washington stressed the dangers and difficulties inherent in postauthoritarian transitions.[31]

Why this permissiveness? Perhaps, in keeping with Schoultz, because Washington did not perceive any direct threat to U.S. security interests. In Central America, of course, the U.S. exercised a major tutelary role. Elections in El Salvador were directly supervised by U.S. personnel. Sim-

[31] See Jeane J. Kirkpatrick, "Dictatorships and Double Standards," *Commentary*, November 1979, republished in idem, *Dictatorships and Double Standards: Rationalism and Reason in Politics* (New York: Simon and Schuster, 1982), 23–52.

ilarly, the transitions in Honduras and Guatemala occurred under the watchful scrutiny of Washington. (Panama stands in a class by itself.) There is a legitimate question about whether or not these cases constitute instances of "democratization," but there is no doubt about the reality of external influence. Basically, the Reagan administration supported these changes in order to consolidate friendly governments and to extend the diplomatic and political isolation of Nicaragua.

The relatively passive stance of the U.S. regarding the transitions South America—in Uruguay, Argentina, Brazil, Peru, and most recent Chile—may stem from two reasons. First, the U.S. has less direct infl ence on these countries (although it can still try, as shown by Chile 1973). Second, and more importantly, none of these transitions threa ened to empower a radical party of the Left. With or without dire supervision of the armed forces, every one of these transformations cu minated in contests between parties of the Center and the Right (the onl possible exception being Brazil in 1989). Washington perceived no ger uine threat to U.S. security interests, so there was no need to intervene

One might imagine that the end of the cold war would augur well fo Schoultz's "liberals." Where are they now that their hour has arrived Ensconced in the bureaucracy, we trust, just as before, silently and sel lessly responding to the demands, incentives, and temptations of offici life in Washington. In fact, *National Security* displays a keen and sym pathetic sense for the situational predicaments of policymakers, an out look suggesting that bureaucratic and political constraints—not person beliefs or attitudes—can determine policy outcomes. Liberals could in habit every nook and cranny of the policy-making apparatus, but i would not make any difference if they were kept under wraps by insti tutional superiors.

From Central America and the Caribbean we turn to the history o U.S. relations with Argentina, to what Joseph Tulchin labels a "conflicted relationship." His chronological coverage is a bit unbalanced, centering on the World War II era, with only cursory attention to the past two decades. Nor is there much of an interpretive argument. Tulchin asserts that the difficulty in U.S.-Argentine relations has structural roots, and here he fastens on the persistence of fantasy within policy-making circles in Buenos Aires: "Argentine leaders aspired to greatness and to influence in world affairs but refused to compete with the other great powers on the usual terms. As a consequence, the Argentine view of world affairs became unrealistic even as it became entrenched in the thinking of the nation's leaders." All in all, Tulchin declares, Argentina's leaders "never had a global sense of their national interest" (p. 25).

One strength of this book is its focus on foreign policy questions from Latin American perspective, a welcome contribution to the appallingly hin literature on Latin American foreign policies.[32] But at times the mphasis on ideological tradition can lead to dubious judgments. Dismissing the Argentine penchant for unyielding adherence to high-ninded principle (*principismo*), for instance, Tulchin ridicules a foreign ninister's statement in the 1980s that his country was seeking to become moral power. "It is not clear what it means to be a moral power or a noral nation," Tulchin remarks. "It is not a familiar category in political cience or in diplomacy" (p. 162). It is not so unfamiliar as a rhetorical actic, however, especially if one is trying to redefine the issues under negotiation. Also curious is Tulchin's insistence that Argentina must ome to terms with the United States: "Until and unless Argentina can esolve its long-standing differences with the United States," he writes, its foreign policy will continue to be characterized by contradictions, a onfrontational tone in the face of minor crises, and a *principismo* that is ittle congruent with other aspects of the international comportment of he nation" (p. 160). But what does a resolution of "long-standing differ-nces" with the U.S. actually entail? Capitulation to the security argu-nents that Schoultz so thoroughly demolished? In fact, Tulchin con-edes that Argentina's stance on Central America was not 'confrontational or quixotic" but "realistic and moderate" (p. 164). In-explicably failing to mention the Plan Austral, he also points out that Argentina's tactics for debt negotiations made a good deal of sense. His only real criticism of the Alfonsín administration concerns an ill-timed rip to Havana (pp. 168–69). It is not quite clear where Argentina went so wrong.

Even so, the Tulchin book makes several significant points. Its histor-cal narrative offers a poignant reminder about the deep roots of antag-onism toward the U.S., not only in countries where we have intervened nilitarily but also in the more distant reaches of South America. It em-phasizes the role of perceptions in policy-making, especially from the Latin American side. It demonstrates that Latin America has held a modest place in U.S. priorities, as shown by Washington's neglect of requests for economic assistance in the wake of World War II—just as the Truman administration was getting ready to launch the Marshall

[32] The standard source in English is still Jennie K. Lincoln and Elizabeth G. Ferris, eds., *The Dynamics of Latin American Foreign Policies: Challenges for the 1980s* (Boulder, Colo.: Westview Press, 1984). Over the past fifteen years Latin American scholars have produced a growing number of foreign policy studies for individual countries; it is now time for a fresh synthesis.

Plan for Europe. And, largely by omission, it suggests that the U.S. has had a negligible role in the promotion of democracy in Argentina.

What now? Now that the cold war is coming to an end, what will U.S. policy be? What should it be? Abraham F. Lowenthal attempts to supply an answer. Longtime executive director of the Inter-American Dialogue and a prominent expert on U.S.-Latin American relations, Lowenthal asserts that Latin America now presents the United States with two fundamental challenges—the resuscitation of economic growth and the consolidation of democracy (undefined!). Writing in the mid-1980s, Lowenthal spells out the possibilities:

> The United States must take the initiative if Latin America's crisis of the 1980s is to be turned to opportunity. The time has come for bold U.S. leadership, not intrusive or domineering but respectful and supportive.
>
> The United States should propose measures to help reexpand the flow of funds, goods, and services throughout the Americas; strengthen democracy and protect human rights; curb the deadly traffic in narcotics; and promote hemispheric peace and security, especially in Central America. On all these major issues, the countries of Latin America and the United States share fundamental objectives. On each, despite considerable differences of interest and perspective, great potential exists for enhanced inter-American collaboration. But new policies are needed to convert compatible aims into effective partnership. (p. 183)

In the economic sphere, Lowenthal argues, the U.S. has a major stake in Latin America's recovery. Since the reduction of debt burdens and an influx of capital will not come about through free-market decisions, governmental action is vital. The U.S. should accordingly concentrate its activity in five areas: supporting international financial institutions, inducing commercial banks to renew lending, proposing a plan to distribute costs of some debt relief, redoubling efforts to reduce the U.S. fiscal deficit and real interest rates, and promoting liberalized trade. In the political realm, Lowenthal urges the U.S. and Latin America to support fundamental human rights; to fortify the social underpinnings of democracy "by confronting the gross inequities in Latin America that breed revolution and repression"; and to support "the institutions and practices that are the fabric of democracy"—including legislatures, civilian control of military, law enforcement, political parties, and the press (p. 188). According to this argument, the United States has the capacity to promote democracy in Latin America but has not done so in the recent past. The time has come for a new U.S. stance, concludes Lowenthal, "one that is built on confidence and trust, on what Latin Americans call *confianza*" (p. 199).

owenthal's prescriptions are workable and wise. Some are
ɔrtatory (it is easier to "confront" gross inequities than to
for instance). Despite its judicious tone, however, much of
ıdvice appears to have fallen on deaf ears. "Democracy can-
:d by force," he insists (p. 194); Washington's post–cold war
: since invaded Panama, asphyxiated Nicaragua, and dis-
ary advisers to the Peruvian highlands. The Reagan-Bush
ns appear to have concluded, understandably if ultimately
ıat economic and military aggression can achieve desired
Outside the diplomatic drawing rooms, talk of *confianza*
nnection to contemporary reality.[33]
ɛ of this book, as for the other studies under review, is the
heoretical framework. The call for *confianza* is an exhorta-
nceptual tool. This is all the more conspicuous since Low-
er the years made such important analytical contributions
standing of U.S. policy-making toward Latin America.[34]
lludes to these issues from time to time, especially in his
:he decline of U.S. hegemony (chap. 2), he offers no overall
ɪr the discussion. As a result his policy prescriptions, valu-
themselves, are not part of a coherent strategy. The whole
ɛ sum of the parts.

Reprise: In Quest of Theory

:al paucity that has come to characterize recent work on
an politics carries considerable costs. This is not to argue
ry is better than no theory. (Even this assertion raises de-
;. First, no analysis can be totally devoid of theory: where
be absent, theory is usually undeveloped and implicit,
ɔ inconsistency. Second, outmoded theory makes a contri-
ng tested and rejected, it allows dismissal of arguments and
understanding.) In fact, as already noted, there are good
ven sufficient reasons for the relative modesty of theory in
ıre: the complexity of political transitions in Latin America
ile generalization.

provement in the tone of U.S.-Latin American relations under the Bush
s more to do with realpolitik than with *confianza*, as Lowenthal himself has
: Abraham F. Lowenthal, "Rediscovering Latin America," *Foreign Affairs*
-41.

:aham F. Lowenthal, "United States Policy toward Latin America: 'Lib-
nd 'Bureaucratic' Perspectives," *Latin American Research Review* 8 (Fall

At issue has been the relationship between economics and politics, particular, the connection between macroeconomic change and the pr cess of regime transformation. Contemporary transitions from authoı tarianism toward democracy in Latin America have occurred und highly inauspicious conditions. This tendency has sharply challenged d pendency theory and apparently refuted its once-fashionable assertic that Latin America faced only two plausible long-term political ou comes, fascism or socialism. It has also created considerable confusic within the field of Latin American politics.

How to explain the emergence of political democracy amid econom crisis? Scholars have sought to resolve this paradox in a variety of way In so doing, they have begun to articulate elements for the constructic of new and viable theory. One approach has been to broaden the notic of causality. There has come forward the hypothesis that economic crisi in the 1930s as in the 1980s, would be likely to undermine *any* kind political regime, democratic or authoritarian. Economic downturn pr cipitates political transition, in this view, but without regard to the *dire tion* of change: it could undermine democratic or authoritarian regime whichever happened to be in place at the time. And by extension, th proposition gives rise to an additional hypothesis: the continuation economic crisis would complicate and endanger the *consolidation* of an ensuing democratic regime. Economic variables thus have considerabl impact on political outcomes. In contrast to both modernization or de pendency frameworks, however, the consequent direction of politic change does not carry specified positive or negative signs.

An alternative outlook asserts that political variables, not economi circumstances, determine political outcomes. There is no necessary o intrinsic relationship between economics and politics; the task is to com prehend the state itself, rather than reduce it to a social artifact. Thi focus leads to an emphasis on statecraft, on elite maneuvering to achiev regime transition within a context of serious constraints. Implicitly, how ever, this orientation also demands attention to the variable roles of civi society, to the appeal of the "democratic idea," and to elite-mass rela tions.[35] And in the process, this focus resuscitates some of the key con cerns of modernization theory. Viewed in retrospect, the modernization school tended to concentrate on the conditions and processes for what would now be called the formation and empowerment of civil society, and this might be its most enduring contribution and its greatest strength.

[35] As emphasized by Levine (fn. 16).

A third attempt to resolve the Latin American paradox concentrates not so much on intranational factors as on external political influences. There exists a contention that the United States has played a major role in the encouragement of political democracy in Latin America (but as the books under review demonstrate, there is little basis for this claim in recent history). A more challenging version of this argument emphasizes the role of such sweeping alterations in the international arena as the acceleration of communications and technology and the extension of economic interdependence. These factors are widely interpreted as key elements in the discrediting of socialist ideologies, the termination of the cold war, and the process of political transition in Eastern Europe.[36] Theoretically, at least, it might seem tempting to interpret the Latin American transitions in a similar light. One problem is that transformations in Latin America occurred well before they took place in Eastern Europe. A second is that state-society relationships and, in particular, the roles of intellectuals seem to have been entirely different. Yet another is that efforts at comparison have thus far been more declarative than analytical, more assertive than insightful. We have urgent need for systematic comparisons of democratization in Latin America with analogous movements in Eastern Europe and other parts of the world.

A final element in the quest for intellectual order derives from the contention that, despite superficial and cosmetic changes, the transformations of the 1980s do not really represent fundamental transitions from authoritarian to democratic regimes. They represent modifications and adaptations of authoritarian patterns, with conspicuous changes in personnel and procedures for executive selection but with underlying continuities in allocations of power and processes of policy-making. The value of the dependent variable (regime type) has not varied as much as is commonly claimed, so there is less variance to explain. In its ultimate form, this argument may even salvage one of the basic premises of the dependency school—that uneven and inequitable socioeconomic development in Latin America would lead to the imposition of nondemocratic regimes. Semiauthoritarian *democraduras* are quite consistent with this general proposition.

All these observations, or hypotheses, tend to function at the intermediate or middle range. None of them purports to offer a grand theory, but they offer guidelines and building blocks. Examination of these propositions serves to emphasize three fundamental points: first, the importance of defining concepts and specifying variables (such as democ-

[36] See Dankwart Rustow, "Democracy: A Global Revolution?" *Foreign Affairs* 69 (Fall 1990), 75–91.

ratization, either as transition or consolidation); second, the need to cor struct explicit models for interrelationships; and third, the utility (complex models that include international as well as domestic factor social and political forces as well as economic determinants. Finally an perhaps most important is a question of intellectual style. Analysts c political change in Latin America would do well to build upon previou contributions, rather than denounce and discard competing theories an theoreticians à outrance. This is not a call for mindless eclecticism; it i an appeal for rigorous attention to empirical reality and to the analytica (rather than ideological) properties of alternative frameworks. Especiall in view of the contemporary complexity of Latin America, it now seems appropriate to construct theories and models gradually—block by block from the bottom up, rather than in top-down fashion.

In the long run the surge and decline of theoretical orthodoxies ha subjected our understanding of Latin America to a pattern of cycles rather than to one of cumulative growth. The confusion of the 1980s ha had a chastening effect on scholarship. Beginning with this modesty, per haps the 1990s will witness a revival of the theoretical vigor, innovation and disputation long characteristic in the field of Latin American studies